PHD

Large deviations and basic information theory for hierarchical and networked data structures

Doku-Amponsah, Kwabena

Award date:
2006

Awarding institution:
University of Bath

Link to publication

LARGE DEVIATIONS AND BASIC INFORMATION THEORY FOR HIERARCHICAL AND NETWORKED DATA STRUCTURES

submitted by

Kwabena Doku-Amponsah

for the degree of Doctor of Philosophy

of the

University of Bath

Department of Mathematical Sciences

August 2006

To my family

Abena, Nana and Papa

Summary

For any finite coloured graph we define the *empirical neighbourhood measure*, which counts the number of vertices of a given colour connected to a given number of vertices of each colour, and the *empirical pair measure*, which counts the number of edges connecting each pair of colours. For a class of models of sparse coloured random graphs, we prove large deviation principles for these empirical measures in the weak topology. The rate functions governing our large deviation principles can be expressed explicitly in terms of relative entropies. We derive a large deviation principle for the degree distribution of Erdős-Rényi graphs near criticality. Using these large deviation principles and others, we prove *asymptotic equipartition properties* for hierarchical structures (modelled as multitype Galton-Watson trees) and networked structures (modelled as coloured random graphs).

Acknowledgments

I wish to express my sincere gratitude to Almighty God for helping me to complete this Thesis. I would also like to express my deepest appreciation to Peter. Mörters for his good supervision, support and fatherly advice. Thank you N. Sidorova for providing the main idea in the proof of Lemma 2.5.5. I wish also to thank S. C. Harris and Peter. Eichelbacher for accepting to examine this Thesis. To my office mates; James Coughlan, Jason Cook, Matt Dorey, Damien Harwin, and Lorina Varvaruca, I would like to say a 'big' thank you for your help and friendship. My profound thanks also goes to the Department of Mathematical Sciences, University of Bath, without whose studentship it would have taken additional time to finish my research. I wish to acknowledge the computer support team, the administrative staff (Jill, Mary, Dena) and the academic staff for their support. Finally, to my wife Ewura Abena I would like to say many thanks for your support and encouragement during hard times.

Contents

Chapter 1

Introduction

Information is often structured in a nonlinear way. For example, in genetics information often has an implicit hierarchical structure, in computer science data is often organized in the form of a network. To transmit or compress data from these sources, one needs efficient coding schemes and approximate pattern matching algorithms, and *large deviations* via the *asymptotic equipartition property* or *Shannon-McMillan-Breiman theorem*, see [CT91], play a key role in this regard, for example by providing bounds on the possible performance of algorithms.

Large deviations is the rate of decay of the probabilities a family of probability measures assign increasingly unlikely events.

In this thesis we investigate the large deviations for a very simple random model of networked data. The data is described as units, chosen from a finite alphabet, together with a random number of links connecting the units. Such an object can be described as a random graph where each vertex of the graph carries a random unit or colour. The easiest model is that of an Erdős-Renyi graph where each vertex is equipped with an independently chosen colour. The slightly more realistic models we consider here allow for a dependence between colour and connectivity of the vertices.

With each coloured graph we associate its *empirical neighbourhood measure*, which records the number of vertices of a given colour with a given number of adjacent vertices of each colour. From this quantity one can derive a host of important characteristics of the coloured graph, like its degree distribution, the number of edges linking two given colours, or the number of isolated vertices of any colour. The first aim of this project is to derive a large deviation principle for the empirical neighbourhood measure.

To be more specific about our model, we consider coloured random graphs constructed as follows: In the first step each of n fixed vertices independently gets a colour, chosen according to some law μ on the finite set $\mathcal{X}$ of colours. In the second step we connect any pair of vertices independently with a probability $p(a, b)$ depending on the colours $a, b \in \mathcal{X}$ of the two vertices. This model, which comprises the simple Erdős-Rényi graph with independent colours as a special case, was introduced by Penman in his thesis [Pe98], see [CP03] for an exposition.

Our first concern in this thesis are asymptotic results when the number n of vertices go to infinity, while the connection probabilities go to zero of order $1/n$. This leads to an average number of edges of order n, the *near-critical* or *sparse* case. Our methods also allow the study of the *sub- and supercritical* regimes. Some results on these cases are discussed in Chapter 3.

Apart from the empirical neighbourhood measure defined above, we also consider the *empirical pair measure*, which counts the number of edges connecting any given pair of colours, and the *empirical colour measure*, which simply counts the number of vertices of any given colour. We prove a joint large deviation principle for the empirical neighbourhood measure and the empirical pair measure of a coloured random graph in the weak topology, see Theorem 2.3.1. In the course of the proof of this principle, two further large deviation principles are established: A large deviation principle for the empirical neighbourhood measure conditioned to have a given empirical pair and colour measure, see Theorem 2.5.1, and a joint large deviation principle for the empirical colour measure and the empirical pair measure, see Theorem 2.4.4. For all these principles we obtain a completely explicit rate function given in terms of relative entropies.

As an example, we consider the Erdős-Rényi graph model on n vertices, where edges are inserted with probability $p_n \in [0, 1]$ independently for any pair of vertices. We assume that $np_n \to c \in (0, 1)$. From our main result we derive a large deviation principle for the degree distribution, see Corollary 2.3.2. This example seems to be new in this explicit form.

Our second concern in the thesis are asymptotic equipartition properties for simple hierarchical and networked structures. We use the large deviation techniques, as provided in the first part of this thesis and [DMS03], to study the asymptotic equipartition property (AEP) of structured data consisting of a large number of *units*, chosen from a finite set, together with a number of *links* connecting the units.

The asymptotic equipartition property (AEP) is a general property used extensively in *information theory* concerning the output samples of a stochastic source. Roughly speaking, it says the *typical output samples of a stochastic data source are all equally likely.* The AEP is a simple consequence of the weak law of large numbers (WLLN). The WLLN for an empirical measure on a finite space is "often a by-product" of a *large deviation principle* (or LDP for short) for the empirical measure.

Two major sets of research work on the AEP (and its applications) within mathematics and information theory have so far been considered. The first of these has focussed on stationary ergodic processes such as Markov chains, see [CT91] and the references therein. The second has dealt with stationary (ergodic) random fields on $\mathbb{Z}^d$, as well as amenable group actions, see, for example [DK02] and the reference therein. Whilst typical examples of applications of the former has concentrated on data from linear source, the latter includes recent advances such as image and video processing, geostatistics, and statistical mechanics.

However, numerous types of data we usually come across in applications (communication studies, demographic studies, biological population studies and the field of physics) are naturally structured like networks or trees. For example, the WWW (consisting of a collection of pages residing on a server with a given name, together with 'hyperlinks' with their direction ignored), data on the spread of some disease in a given population and many more, can be described by networks. Equally, the age structure of a given population is best modelled by trees. As an application of our abstract principles, we consider the following concrete examples from biology.

- **Metabolic network:** This is a graph of interactions forming a part of the energy generation and biosynthesis metabolism of the bacterium E.coli. Here, the units represent *substrates* and *products*, and links represent *interactions*. See [New00].

- **Mutation study:** Consider mutations in mitochondrial DNA (mtDNA for short) e.g. the mtDNA^{4977} deletion (a mutation which causes a deletion of about one third of the mitochondrial genome). The replication of mtDNA can be described by a tree, where the units are *normal* and *mutant* and the links indicate 'mother-child' relations. See [OS02] and the references therein.

The core results of the second half of this thesis are the *Shannon-McMillan-Breiman theorems* for two simple probabilistic models: The *multitype Galton-Watson trees* describing hierarchical data structures, and a class of *coloured random graphs* describing networked data structures, see Theorems 3.4.1 and 3.2.1.

Specifically, we consider for the first model typed trees described by the following procedure: The root carries random type chosen according to the some law on a finite alphabet; given the type of a vertex, the number and types of the children (ordered from left to right) are given independently of everything else, by an offspring law. We review this model in the next section, Section 1.1.

Finally, our results can also be used to understand simple models of *statistical mechanics* defined on random graphs. Here, typically, the colours of the vertices are interpreted as spins, taken from a finite set of possibilities, and the Hamiltonian of the system is an integral of some function with respect to the empirical neighbourhood measure. As a very simple example we provide the annealed asymptotics of the random partition function for the Ising model on an Erdős-Rényi graph, as the graph size goes to infinity. See Chapter 2, Example 2.

The rest of this thesis is organized in the following way. In Section 1.1 we review the model for hierarchical structure, see [DMS03], for the multitype Galton-Watson trees. We review the model for networked structures, coloured random graph and extend some known concepts to this models in Section 1.2. The penultimate section contains a review of the method of mixtures in large deviations, taken from [Bi04]. We end the chapter by providing an overview of the main chapters. See Section 1.4.

In Chapter 2 we present the LDP for the empirical neighbourhood measure and the empirical pair measure on near-critical coloured graphs. Major tools used are (exponential) change of measure, the method of types and the method of mixtures.

We discuss in Chapter 3 the *asymptotic equipartition properties* for structured data modelled as *either* multitype Galton-Watson tree *or* coloured random graphs. As examples, we compute the asymptotic number of bits needed to encode large amount of data from the model of the mtDNA^{4977} as well as the metabolic network of the bacterium E. Coli.

Recall that a *rate function* is a non-constant, lower semicontinuous function I from a polish space $\mathcal{M}$ into $[0,\infty]$, it is called *good* if the level sets $\{I(m) \le x\}$ are compact for every $x \in [0,\infty)$.

A functional U from the set of finite graphs to $\mathcal{M}$ or the distribution of U $P_n(\,\cdot\,) = \mathbb{P}\{\,U(X) \in \cdot\,\}$ is said to satisfy a *large deviation principle* with speed b_n ($b_n \to \infty$) and rate function I if, for all Borel sets $B \subset \mathcal{M}$,

$$-\inf_{m\in\operatorname{int} B} I(m) \le \liminf_{n\to\infty} \tfrac{1}{b\,n} \log P_n(B) \le \limsup_{n\to\infty} \tfrac{1}{b\,n} \log P_n(B) \le -\inf_{m\in\operatorname{cl} B} I(m)\,,$$

where X under $\mathbb{P}$ is a random graph with n vertices, and int B and cl B refer to the interior, resp. closure, of the set B. This Thesis considers both cases when $b_n = n$ and $b_n = a_n n^2$, for $a_n \to 0$. See Chapters 2 and 3.

1.1 Simple Hierarchical Structures

We review in this section the model for simple hierarchical data structures, multitype Galton-Watson trees as presented in [DMS03].

Notation: Denote by $\mathcal{T}$ the set of all finite rooted planar tree T, by $V = V(T)$ the set of all vertices and by $E = E(T)$ the set of all edges oriented away from the root, which is always denoted by ρ. Write $|T|$ for the number of vertices in the tree T. Let $\mathcal{X}$ be a finite alphabet.

Suppose that T is any finite tree and we are given an initial probability measure μ on a finite alphabet $\mathcal{X}$ and a Markovian transition kernel $Q\colon \mathcal{X} \times \mathcal{X} \to [0,1]$. We can obtain a *tree indexed Markov chain* $X : V \to \mathcal{X}$ by choosing $X(\rho)$ according to μ and choosing $X(v)$, for each vertex $v \ne \rho$, using the transition kernel given the value of its parent, independently of everything else. If the tree is chosen randomly, we always consider $X = (X(v) : v \in T)$ under the *joint law* of tree and chain. It is sometimes convenient to interpret X as a *typed tree*, considering $X(v)$ as the *type* of the vertex v.

We write

$$\mathcal{X}^* = \bigcup_{n=0}^{\infty} \{n\} \times \mathcal{X}^n$$

and equip it with the largest topology with all subsets as open sets (discrete topology). We shall interpret $\mathcal{X}^*$ as the space of numbers and types of immediate offspring of a vertex.

1.1.1 The Multitype Galton- Watson Model

We denote by N random number of offspring particles and by $X_1, ..., X_N$, N random offspring particles. Let μ be a probability measure (initial distribution) on $\mathcal{X}$ and $\mathbb{Q}$ an offspring transition kernel from $\mathcal{X}$ to $\mathcal{X}^*$. We define the law $\mathbb{P}$ of a tree-indexed process X by the following procedure:

- The root ρ carries a random type $X(\rho)$ chosen according to the probability measure μ on $\mathcal{X}$;
- for every vertex with type $a \in \mathcal{X}$ the offspring number and types are given independently of everything else, by the offspring law $\mathbb{Q}\{\cdot\,|\,a\}$ on $\mathcal{X}^*$. We write
$$\mathbb{Q}\{\,\cdot\,|\,a\} = \mathbb{Q}\{(N, X_1, ..., X_N) \in \,\cdot\,|\,a\}.$$

We always consider $X = ((X(v),\, C(v)),\, v \in V)$ under the joint law of tree and offspring. We interpret X as multitype Galton-Watson tree and $X(v)$ as the type of vertex v. Notice the offspring of any vertex $v \in T$ is characterized by an element of $\mathcal{X}^*$ and that there is an element $(0, \emptyset)$ in $\mathcal{X}^*$ symbolizing absence of offspring.

For each typed tree X and each vertex v we denote by

$$C(v) = (N(v), X_1(v), \ldots, X_{N(v)}(v)) \in \mathcal{X}^*,$$

the number and types of the children of v, ordered from left to right.

Definition 1.1.1. *We call an offspring kernel $\mathbb{Q}$ bounded if for some $N_0 < \infty$,*

$$\mathbb{Q}\{N > N_0\,|\,a\} = 0, \;\; \forall\, a \in \mathcal{X}\,.$$

Definition 1.1.2 (Multiplicity). *Denote, for every $c = (n,\, a_1, \ldots, a_n) \in \mathcal{X}^*$ and $a \in \mathcal{X}$, the multiplicity of the symbol a in c by*

$$m(a, c) := \sum_{i=1}^{n} 1_{\{a_i = a\}}.$$

Definition 1.1.3. *Define the matrix A with index set $\mathcal{X} \times \mathcal{X}$ and non-negative entries by*

$$A(a, b) = \sum_{c \in \mathcal{X}^*} \mathbb{Q}\{c\,|\,b\} m(a, c), \;\; \textit{for } a, b \in \mathcal{X}.$$

$A(a, b)$ is the expected number of offspring of type a of a vertex of type b.

Definition 1.1.4. $A^*(a,b) := \sum_{k=1}^{\infty} A^k(a,b) \in [0,\infty]$, *for* $a, b \in \mathcal{X}$.

Definition 1.1.5 (Irreducibility). *The matrix A is irreducible if*

$$A^*(a,b) > 0, \; \forall a, b \in \mathcal{X}.$$

Definition 1.1.6. *The multitype Galton-Watson tree is called irreducible if the matrix A is irreducible.*

Definition 1.1.7 (Criticality). *It is called critical (subcritical, supercritical) if the largest eigenvalue of the matrix A, λ_A is* 1 *(less than* 1*, greater* 1 *respectively).*

We restrict our attention to models of critical multitype Galton-Watson tree.

Proposition 1.1.8. *[DZ98, Theorem 3.1.1]. If π is the eigenvector of the matrix A corresponding to λ_A (normalized to a probability vector), then, π is unique.*

Definition 1.1.9. *We define the invariant distribution $\pi \otimes \mathbb{Q}(\cdot,\cdot)$ of the multitype Galton-Watson tree X by*

$$\pi \otimes \mathbb{Q}(a,\, c) = \pi(a)\mathbb{Q}\{c\,|\,a\}, \; \textit{for} \; (a,c) \in \mathcal{X} \times \mathcal{X}^*.$$

Definition 1.1.10 (Empirical offspring measure). *We define for every multitype Galton-Watson tree $X = ((X(v), C(v)),\, v \in V)$ the empirical offspring measure $M_X \in \mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$ by*

$$M_X(a,c) = \frac{1}{|T|} \sum_{v \in V} \delta_{(X(v),C(v))}(a,c), \quad \textit{for} \; (a,c) \in \mathcal{X} \times \mathcal{X}^*.$$

We denote by ν_1 the $\mathcal{X}-$ marginal of the probability measure ν.

Definition 1.1.11. *We call ν shift-invariant if*

$$\nu_1(a) = \sum_{(b,c) \in \mathcal{X} \times \mathcal{X}^*} m(a,c)\nu(b,c), \;\; \forall a \in \mathcal{X} \; . \tag{1.1.1}$$

If ν is an empirical offspring measure of multitype Galton-Watson tree, then both sides of (1.1.1) is

$$\tfrac{1}{|T|}\sharp\{\text{number of vertices of type } a\}.$$

By the concept of shift-invariant and the recent result LDP for empirical offspring measure we prove WLLN for M_X, and use it to obtain the AEP for hierarchical structures. See Chapter 3.

1.2 Simple Networked Structures

In this subsection, we review the model for simple networked structures, *coloured random graph model.*

Notation: We begin by fixing the following notations. Let V be a fixed set of n vertices, say $V = \{1, \ldots, n\}$. Denote by $\mathcal{G}_n$ the set of all (simple) graphs with vertex set V and edge set $E \subset \mathcal{E} := \big\{(u, v) \in V \times V \,:\, u < v\big\}$, where the *formal* ordering of edges is introduced as a means to simply describe *unordered* edges. Note that for all n, we have

$$0 \leq |E| \leq N(n) = \tfrac{n(n-1)}{2}.$$

Let $\mathcal{X}$ be a finite alphabet or colour set.

1.2.1 The Coloured Random Graph Model

Given a symmetric function $p_n\colon \mathcal{X} \times \mathcal{X} \to [0, 1]$ and a probability measure μ on $\mathcal{X}$ we may define the *randomly coloured random graph* or simply *coloured random graph* X with n vertices as follows:

- Assign to each vertex $v \in V$ colour $X(v)$ independently according to the *colour law* μ.

- Given the colours, we connect any two vertices $u, v \in V$, independently of everything else, with *connection probability*

$$p_n(X(u), X(v)).$$

We always consider $X = ((X(v) \,:\, v \in V), E)$ under the joint law of graph and colour and interpret X as coloured random graph.

We denote by $\mathbb{P}$ the law of coloured random graph X with n vertices.

Denote by $\mathcal{G}_n(\mathcal{X})$ the set of all coloured graphs with colour set $\mathcal{X}$ and n vertices.

1.2.2 Types of Coloured Random Graphs

Definition 1.2.1. *We call X sparse or near-critical if the connection probabilities satisfy*

$$a_n^{-1}p_n(a,b) \to C(a,b), \qquad \forall a,b \in \mathcal{X},$$

where $C\colon \mathcal{X}\times\mathcal{X} \to [0,\infty)$ is a nonzero function and $a_n n \to 1$.

Definition 1.2.2. *We call X subcritical if the connection probabilities satisfy*

$$a_n^{-1}p_n(a,b) \to C(a,b), \qquad \forall a,b \in \mathcal{X},$$

where $C\colon \mathcal{X}\times\mathcal{X} \to [0,\infty)$ is a nonzero function and $a_n n \to 0$.

Definition 1.2.3. *We call X dense or supercritical if the connection probabilities satisfy*

$$a_n^{-1}p_n(a,b) \to C(a,b), \qquad \forall a,b \in \mathcal{X},$$

where $C\colon \mathcal{X}\times\mathcal{X} \to [0,\infty)$ is a nonzero function and $a_n n \to \infty$.

Assumptions: We assume throughout the thesis that $C\colon \mathcal{X}\times\mathcal{X} \to [0,\infty)$ is not identically equal to zero. We also assume that the sequence (a_n) converges to 0 as n approaches infinity.

1.2.3 Empirical Measures on Coloured Random Graphs

Notation: For any finite or countable set $\mathcal{Y}$, we denote by $\mathcal{N}(\mathcal{Y})$ the space of counting measures on $\mathcal{Y}$, i.e. those measures taking values in $\mathbb{N}\cup\{0\}$, endowed with the discrete topology, by $\mathcal{M}(\mathcal{Y})$ the space of probability measures, and by $\tilde{\mathcal{M}}(\mathcal{Y})$ the space of finite measures on $\mathcal{Y}$, both endowed with the weak topology. By $\tilde{\mathcal{M}}_*(\mathcal{Y}\times\mathcal{Y})$ we denote the subspace of symmetric measures in $\tilde{\mathcal{M}}(\mathcal{Y}\times\mathcal{Y})$ and by $\|\varpi\|$ we denote the total mass of the finite measure ϖ.

Definition 1.2.4. *Define for every coloured random graph X, the empirical colour measure $L^1 \in \mathcal{M}(\mathcal{X})$ by,*

$$L^1(a) = \frac{1}{n}\sum_{v\in V}\delta_{X(v)}(a), \quad \textit{for } a \in \mathcal{X}. \tag{1.2.1}$$

$L^1(a) = \frac{1}{n}\sharp\{\text{number of vertices of colour } a\}.$

Definition 1.2.5. *For every coloured random graph X, the empirical pair measure $L^2 \in \tilde{\mathcal{M}}_*(\mathcal{Y} \times \mathcal{Y})$ is defined by,*

$$L^2(a,b) = \frac{1}{a_n n^2} \sum_{(u,v)\in E} [\delta_{(X(v),\,X(u))} + \delta_{(X(u),\,X(v))}](a,b), \quad \textit{for } a,b \in \mathcal{X}. \tag{1.2.2}$$

Observe that $a_n n^2$ is some expected number of edges in the graph, and notice from (1.2.2) that

$$L^2(a,b) = \tfrac{1}{a_n n^2}(1 + \mathbb{1}_{\{a=b\}})\sharp\{\text{ edges between vertices of colours } a \text{ and } b\,\}.$$

We look at the LDP for the (L^1, L^2), on the scale n, for coloured random graphs with $a_n^{-1} = n$ in Chapter 2. By exponential equivalence, see [DZ98, Theorem 4.2.13], one can obtain the same LDP for any sparse coloured graphs.

Definition 1.2.6. *We define for every coloured random graph X, the empirical neighbourhood measure $M \in \mathcal{M}(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$ by,*

$$M(a,\ell) = \frac{1}{n}\sum_{v\in V} \delta_{(X(v),\,L(v))}(a,\ell), \quad \textit{for } (a,\ell) \in \mathcal{X} \times \mathcal{N}(\mathcal{X}), \tag{1.2.3}$$

where $L(v) = (l^v(b) \colon b \in \mathcal{X})$ and $l^v(b)$ is the number of vertices of colour b connected to vertex v.

Definition 1.2.7. *The degree distribution $D \in \mathcal{M}(\mathbb{N} \cup \{0\})$ is given by*

$$D(k) = \sum_{a\in\mathcal{X}} \sum_{\ell\in\mathcal{N}(\mathcal{X})} \delta_k\big(\textstyle\sum_{b\in\mathcal{X}} \ell(b)\big)\, M(a,\ell), \quad \textit{for } k \in \mathbb{N} \cup \{0\}.$$

Remark 1 Note that L^1, M, and D are probability measures, but $\|L^2\| = \frac{2|E|}{a_n n^2}$.

1.3 Method of Mixtures

We review the large deviations for mixtures by Biggins. See [Bi04].

To fix some notation we introduce two more spaces of measures.

Definition 1.3.1. *Define for any natural number n the set* $\mathcal{M}_n(\mathcal{X}) \subset \mathcal{M}(\mathcal{X})$ *by*

$$\mathcal{M}_n(\mathcal{X}) = \left\{ \omega \in \mathcal{M}(\mathcal{X}) \,:\, n\omega(a) \in \mathbb{N},\ \forall a \in \mathcal{X} \right\}.$$

Definition 1.3.2. *We define for any natural number n the space* $\tilde{\mathcal{M}}_n(\mathcal{X} \times \mathcal{X})$ *by*

$$\tilde{\mathcal{M}}_n(\mathcal{X} \times \mathcal{X}) = \left\{ \varpi \in \tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X}) \,:\, \tfrac{n}{1+\mathbb{1}\{a=b\}} \varpi(a,b) \in \mathbb{N},\ \forall a, b \in \mathcal{X} \right\}.$$

We write $\Theta := \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X})$ and $\Theta_n := \mathcal{M}_n(\mathcal{X}) \times \tilde{\mathcal{M}}_n(\mathcal{X} \times \mathcal{X})$.

Let $P^{(n)}$ be a mixing probability measure on the space Θ, concentrated on Θ_n. For each $\theta \in \Theta_n$, let $P_\theta^{(n)}$ be a probability measure on $\mathcal{M}(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$ for which the mapping $\theta \mapsto P_\theta^n(A)$ is measurable on Θ_n for every measurable $A \subset \mathcal{M}(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$.

Definition 1.3.3. *A joint distribution* $\tilde{P}^n$ *obtained by mixing over* θ *is given by*

$$d\tilde{P}^n(\theta,\, x) = dP_\theta^{(n)}(x)dP^{(n)}(\theta). \tag{1.3.1}$$

For our purpose we present in a weak form the large deviation principle for $\tilde{P}^n$ in the paper by Biggins [Bi04, Theorem 5(b)].

To start we write

$$Y := \mathcal{M}(\mathcal{X} \times \mathcal{N}(\mathcal{X})).$$

Theorem 1.3.4 ([Bi04])**.** *Suppose* $P^{(n)}$ *satisfies an LDP with good rate function* ψ *and whenever* $\theta(n)$ *converges to* $\theta \in \Theta$, $P_{\theta(n)}^{(n)}$ *satisfies an LDP with rate* λ_θ. *Let* $\lambda_\theta(y)$ *be jointly lower semicontinuous in* $(\theta,\, y) \in \Theta \times Y$, *and suppose in addition,* $\tilde{P}^{(n)}$ *is exponentially tight. Then,* $\tilde{P}^{(n)}$ *satisfies an LDP with good rate function*

$$\lambda(\theta,\, y) = \psi(\theta) + \lambda_\theta(y).$$

The method of mixtures is part of the 'machinery' used to prove the joint large deviation principle for the empirical pair measure and the empirical neighbourhood measure, Theorem 2.4.4. See Chapter 2.

To be more specific about the role of this set up in our proof, we observe that by construction the law $\mathbb{P}$ of a coloured random graph is only a function of the empirical colour measure L^1 and empirical pair measure L^2. Further, the relationship between the empirical neighbourhood measure M and the pair $(L^1,\, L^2)$ is not one-to-one, that is to say, one could have more than one empirical neighbourhood measures with the same empirical colour measure and empirical pair measure. Consider the following example:

Let $\mathcal{X} = \{a,\, b\}$ and consider empirical neighbourhood measures $\nu, \hat{\nu}$ given by

$$\nu(a,\,(1,0)) = \tfrac{2}{6},\ \nu(a,\,(0,1)) = \tfrac{1}{6},\ \nu(b,\,(1,0)) = \tfrac{1}{6},\ \nu(b,\,(0,0)) = \tfrac{2}{6} \text{ and}$$

$\hat{\nu}(a,\,(1,1)) = \frac{1}{6},\ \hat{\nu}(a,\,(1,0)) = \frac{1}{6},\ \hat{\nu}(a,\,(0,0)) = \frac{1}{6},\ \hat{\nu}(b,\,(0,0)) = \frac{2}{6},\ \hat{\nu}(b,\,(1,0)) = \frac{1}{6}.$
Then, the empirical pair measure corresponding to ν is

$$\begin{aligned}
\sum_{\ell\in\mathcal{N}(\mathcal{X})} \ell(b)\nu(a,\,\ell) &= 0\times\tfrac{2}{6} + 1\times\tfrac{1}{6} = \tfrac{1}{6},\\
\sum_{\ell\in\mathcal{N}(\mathcal{X})} \ell(a)\nu(b,\,\ell) &= 1\times\tfrac{1}{6} + 0\times\tfrac{2}{6} = \tfrac{1}{6},\\
\sum_{\ell\in\mathcal{N}(\mathcal{X})} \ell(a)\nu(a,\,\ell) &= 1\times\tfrac{2}{6} + 0\times\tfrac{1}{6} = \tfrac{2}{6},\\
\sum_{\ell\in\mathcal{N}(\mathcal{X})} \ell(b)\nu(b,\,\ell) &= 0\times\tfrac{2}{6} + 0\times\tfrac{1}{6} = 0.
\end{aligned}$$

Similar computation gives the empirical pair measure corresponding to $\hat{\nu}$ as

$$\begin{aligned}
\sum_{\ell\in\mathcal{N}(\mathcal{X})} \ell(b)\hat{\nu}(a,\,\ell) &= 1\times\tfrac{1}{6} + 0\times\tfrac{1}{6} + 0\times\tfrac{1}{6} = \tfrac{1}{6},\\
\sum_{\ell\in\mathcal{N}(\mathcal{X})} \ell(a)\hat{\nu}(b,\,\ell) &= 0\times\tfrac{2}{6} + 1\times\tfrac{1}{6} = \tfrac{1}{6},\\
\sum_{\ell\in\mathcal{N}(\mathcal{X})} \ell(a)\hat{\nu}(a,\,\ell) &= 1\times\tfrac{1}{6} + 1\times\tfrac{1}{6} + 0\times\tfrac{1}{6} = \tfrac{2}{6},\\
\sum_{\ell\in\mathcal{N}(\mathcal{X})} \ell(b)\hat{\nu}(b,\,\ell) &= 0\times\tfrac{2}{6} + 0\times\tfrac{1}{6} = 0.
\end{aligned}$$

It not hard to see that both $\nu, \hat{\nu}$ have empirical colour measure ω given by

$$\omega(a) = \omega(b) = \tfrac{1}{2}.$$

To get 'round' this problem we fix the empirical colour measure and the empirical pair measure. Fixing the pair (L^1, L^2) leads to a simple model which allowed combinatorial arguments to be used in the proof of the LDP for the empirical neighbourhood measure. We use change of measure technique to get the LDP for (L^1, L^2) and then, use Theorem 1.3.4, to 'mix' the two LDPs to established the joint LDP for the empirical pair measure and the empirical neighbourhood measure.

1.4 Overview of the Main Chapters

1.4.1 Chapter 2: The LDPs for the Empirical Measures of Sparse Coloured Random Graph

Roughly speaking, the empirical neighbourhood measure is a probability measure which records the colour of a vertex and how many neighbours of this vertex have a given colour in a coloured random graph.

Our main concern in this chapter is the LDP for the empirical neighbourhood measure of sparse coloured random graph. See Theorem 2.3.1. To get this result we consider two further large deviation principles.

Firstly, we consider the empirical pair measure (a finite symmetric measure) L^2, which records the number of edges connecting every pair of colours, and the empirical colour measure (a probability measure) L^1, which counts the number of vertices of each colour. For these empirical measures we prove a joint large deviation principle in the weak topology, see Section 2.4.

To prove this principle, the law of the coloured random graph $\mathbb{P}$ is changed on the 'levels' of colour and edge to get a new law $\tilde{\mathbb{P}}$ of a coloured random graph. Thus, given a finite symmetric measure ϖ and a probability measure ω the change is done in such away that typical graphs under $\tilde{\mathbb{P}}$ have approximately the empirical pair measure ϖ and ω. Using, $\tilde{\mathbb{P}}$ we obtain a large deviation upper bound for (L^1, L^2) in a variational formulation. We solve the variational problem, identify the rate function and show that it is a good rate function.

The lower bound is obtained from the upper bound with the $\tilde{\mathbb{P}}$ taking the place of $\mathbb{P}$. We then use the properties of the rate function to show that under $\tilde{\mathbb{P}}$ the probability that $(L^1,\, L^2)$ is not in a small 'rectangle' around $(\omega,\, \varpi)$ asymptotically vanishes with the size of the graph to end the proof of the LDP for $(L^1,\, L^2)$.

Secondly, we prove an LDP for the empirical neighbourhood measure under the law of a new coloured random model obtained by fixing the L^1 and L^2. See Section 2.5.2.

To illustrate a bit the model, we let $(\omega_n,\, \varpi_n) \in \mathcal{M}_n(\mathcal{X}) \times \mathcal{M}_n(\mathcal{X} \times \mathcal{X})$ and take

$$(L^1,\, L^2) = (\omega_n, \varpi_n).$$

We construct from (ω_n, ϖ_n) a new coloured random graph with law $\mathbb{P}_{(\omega_n,\, \varpi_n)}$ described in the following manner.

- Assign colours to the vertices by sampling without replacement from the collection of n colours, which contains any colour $a \in \mathcal{X}$ exactly $n\omega_n(a)$ times;
- for every unordered pair $\{a, b\}$ of colours create exactly $n(a, b)$ edges by sampling without replacement from the pool of possible edges connecting vertices of colour a and b, where

$$n(a,b) := \begin{cases} n\, \varpi_n(a,b) & \text{if } a \neq b \\ \frac{n}{2}\, \varpi_n(a,b) & \text{if } a = b\,. \end{cases}$$

For our proof we introduce a numbering system, which specifies, for each $\{a, b\}$, the order in which edges are drawn in the second step. More precisely, the edge-number k is attached to both vertices connecting the k^{th} edge.

Denote by $\mathcal{K}^{(n)}(\omega_n, \varpi_n)$ the set of empirical neighbourhood measures of graphs with n vertices, empirical colour measure ω_n and empirical pair measure ϖ_n.

To begin we bound the number of measures in $\mathcal{K}^{(n)}(\omega_n, \varpi_n)$ using the technique of integer partition. See Subsection 2.5.3.

With this bound and a combinatorial argument, for any ν_n we compute the large deviation 'upper' probability for the event $\{ M = \nu_n \}$ under $\mathbb{P}_{(\omega_n, \varpi_n)}$ in Subsection 2.5.4.

For a large deviation 'lower' probability for the event $\{ M = \nu_n \}$ of any $\nu_n \in \mathcal{K}^{(n)}(\omega_n, \varpi_n)$ under $\mathbb{P}_{(\omega_n, \varpi_n)}$, we further perform several estimations and approximations. First, we obtain an upper bound on the support of an empirical neighbourhood measure using simple counting technique via mathematical induction. See Subsection 2.5.5.

Then, we approximate in the weak topology a given empirical neighborhood measure by $\nu_n \in \mathcal{K}^{(n)}(\omega_n, \varpi_n)$ with the 'extra' feature that the degree of any vertex is bounded by $n^{1/3}$. This approximation is carried out in *four* steps, in four separate Lemmas in Subsection 2.5.6

We again use a combinatorial argument and the four approximation lemmas to get a large deviation lower probability for the event $\{ M = \nu_n \}$ in our new model. See Section 2.5.7.

The second auxiliary principle, see Theorem 2.5.1, is derived from these large deviation probabilities.

We then use the method of mixtures to 'mix' the first auxiliary principle, see Theorem 2.4.4, and the second auxiliary principle, see Theorem 2.5.1, to obtain the LDP for (L^2, M), see Theorem 2.3.1.

Remark: For any finite graph L^2 and M are related by a simple formula which allows one to compute L^2 from M. As this formula is discontinuous in the weak topology the relationship breaks down in the limit. To obtain a separate LDP for M or L^2 we need to form a contraction and the resulting rate function becomes much less explicit. Similarly, the rate function for L^2 obtained by contracting the LDP for (L^1, L^2) would not be explicit in general. However, in some special cases we can get explicit rate functions by projecting our LDPs.

These cases are given as corollaries in this thesis.

1.4.2 Chapter 3: The Asymptotic Equipartition Properties for Simple Hierarchical and Networked Structures

The AEP is the version of the strong law of large numbers in information theory. It is fundamental to the concept of typical set used in theories of compression. It can be obtained from the WLLN for a carefully defined empirical measure on a stochastic data source.

Our main aim in this chapter is to derive the AEP for structured data source. We model the data source as either multitype Galton-Watson tree or coloured random graph. The AEP for simple hierarchical structures is discussed in Section 3.2. We state the main result, see Theorem 3.2.1, in the first subsection of the section. This result is then applied to a concrete example from biology in the last subsection. Our proof uses the recent large deviations technique for trees presented in [DMS03] and recalled in Section 3.

To be more elaborative about this technique, first we derive from the LDP for the empirical offspring measure M_X of multitype Galton-Watson tree, see Theorem 3.3.1, and the Perron Frobenius theorem, see, for example [DZ98, Theorem 3.1.1], a weak law of large numbers for empirical offspring measure of trees with bounded offspring distribution $\mathbb{Q}$. Next, we compute the distribution $P_n(x)$ of a typed tree x conditioned on the event $\{|T| = n\}$, and express it as a function of the M_x on the tree. Then, by boundedness of the offspring kernel $\mathbb{Q}$, Lemma 3.3.4 and our WLLN for M_x we show that $\langle M_x, -\log \mathbb{Q}\rangle$ converges to

$$\langle \pi \otimes \mathbb{Q}, -\log \mathbb{Q}\rangle,$$

as n approaches infinity to end the proof.

We treat separately the AEP for the networked structures in Section 3.4.1. The main theorem, see Theorem 3.4.1, of the section is AEP for random networks in subcritical, near-critical and supercritical regimes. For random network consisting of n units connected by an average number of order $\log n/n$ links we obtain a most interesting result, see Theorem 3.4.2. To conclude the section we apply Theorem 3.4.1 to a specific example from biology. See Subsection 3.4.2.

The proofs of these AEPs carried out in the last section of the chapter also uses the technique of large deviations. To start, we obtain joint LDPs for empirical colour measure L^1 and the empirical pair measure L^2 of subcritical and supercritical coloured random graphs in the weak topology. These LDPs are formulated in Theorem 3.5.1 and Theorem 3.5.2, respectively. See Subsection 3.5.2. We consider the LDPs on two different scales $a_n n^2$ and n. On both scales we use the technique of change of measure to establish the LDPs. See Subsections 3.5.4 and 3.5.3.

Heuristically, for the LDPs on the scale $a_n n^2$ we change the colour law in a way that it is asymptotically the typical one. That is to say, we change the measure to avoid 'colour cost' or we change the colour law 'for free', but incur some cost on connection. Conversely, for the LDPs on the scale n we change the measure to avoid 'connection cost'. i.e. we change the edge law in a manner that, asymptotically, once a colour law ω is fixed the edge law is the typical one

$$C\omega \otimes \omega.$$

As in the proof of the AEP for hierarchical structures, we derive from our LDPs, Theorem 2.4.4 and Theorem 3.5.1, weak laws of large numbers for L^1 and L^2. See Subsection 3.5.7. Specifically, for a coloured graph $x \in \mathcal{G}_n(\mathcal{X})$ we compute the distribution $P_n(x)$ and express it in terms of the pair

$$(L^1(x),\, L^2(x)).$$

Then, using the conditions on the sequence (a_n), see Theorem 3.4.1, and the weak laws for $L^1(x)$, $L^2(x)$ we show that both $\langle L^1 \otimes L^1,\, -n\log(1-p_n)\rangle$ and $\langle L^2,\, -(\log(p_n/(1-p_n)))/\log n\rangle$ converges to $\|C\mu \otimes \mu\|$, and that $\langle L^1,\, -\log\mu\rangle$ converges to

$$\langle \mu(a),\, -\log\mu\rangle,$$

as n approaches infinity, to prove Theorem 3.4.1 and Theorem 3.4.2.

Chapter 2

The Large Deviation Principles for Empirical Measures of Sparse Coloured Random Graphs

(This material also appears in a co-authored paper with P. Mörters, Bath).

2.1 Introduction

This chapter is devoted to the large deviations study of the empirical neighbourhood measure of near-critical coloured graph model in the weak topology. The study is divided into two main large deviations principles.

In the first, we consider the LDP for (L^1, L^2). The main tool used is the technique of change of measure. By this technique we obtain the upper bound in a variation formulation. We solve the variation problem by using simple technique of identifying rate functions from other rates. See, for example [DZ98, Lemma 6.2.16].

The second is the LDP for the empirical neighbourhood measure conditioned to have a given (L^1, L^2). The proof of this principle is based on combinatorial argument (via the method of types) backed by series of approximations. We control the number of empirical neighbourhood measures by counting integer partitions and bound the support of empirical neighbourhood measures using a simple counting technique.

Using the method of mixtures we combine the auxiliary principles to obtain a joint large deviation principle for (L^2, M) in the weak topology. Using the contraction principle we derive the LDP for M. In the special case of the Erdős-Rényi graphs the rate function looks simple.

The chapter is outlined as follows. Firstly, Section 2.2 contains some notation, definitions and concepts necessary to present the principles (main and auxiliary) in the chapter. To be specific, we introduce the notion of "Consistency" for empirical neighbourhood measures, and highlight a functional relationship between the pair (L^1, L^2) and M, and its effect on our rate functions.

The main LDPs of the chapter, Theorem 2.3.1 and Corollary 2.3.2, are stated in Section 2.3. This corollary is the large deviation principle for the degree distribution D of the Erdős-Rényi graph model.

Section 2.4 contains the LDP for (L^1, L^2), see Theorem 2.4.4, and the proof of this large deviation principle.

In Section 2.5, we state an LDP for the empirical neighbourhood measure for the coloured graph model obtained by fixing the pair (L^1, L^2). This result, see Theorem 2.5.1, is an important step in the proof of Theorem 2.3.1.

The remainder of the chapter contains the proofs of the results set out in Subsection 2.5.1. Unlike the proof of the first auxiliary principle which uses the technique of change of measure, the proof of Theorem 2.5.1, carried out in Subsection 2.5.2, is based on combinatorial arguments combined with a probabilistic approximation technique.

In Section 2.6 we combine these results to obtain our main result, Theorem 2.3.1, using the setup and result of Biggins, see Theorem 1.3.4, to 'mix' the large deviation principles.

Finally, we derive from our main LDP Corollary 2.3.2 in Section 2.7.

2.2 Preparation

We recall $\mathcal{M}(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$ is the space of all probability measures on $\mathcal{X} \times \mathcal{N}(\mathcal{X})$ equipped with the smallest topology which makes the functionals $\nu \mapsto \int f(a,\ell)\,\nu(da\,,d\ell)$ continuous, for $f : \mathcal{X} \times \mathcal{N}(\mathcal{X}) \to \mathbb{R}$ bounded (i.e. the weak topology) and by ν_1, ν_2 the $\mathcal{X}$-marginal and the $\mathcal{N}(\mathcal{X})-$ marginal of the measure $\nu \in \mathcal{M}(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$.

Definition 2.2.1.

$$\langle \nu(\cdot,\ell)\,,\,\ell(\cdot)\rangle(a,b) := \sum_{\ell \in \mathcal{N}(\mathcal{X})} \nu(a,\ell)\ell(b).$$

Definition 2.2.2 (Φ). *The function* $\Phi\colon \mathcal{M}(\mathcal{X} \times \mathcal{N}(\mathcal{X})) \to \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}(\mathcal{X} \times \mathcal{X})$ *is defined by*

$$\Phi(\nu) = (\Phi_1(\nu),\, \Phi_2(\nu)) = (\nu_1, \langle \nu(\cdot,\ell)\,,\,\ell(\cdot)\rangle).$$

Observe that $\Phi(M) = (L^1, L^2)$, if these quantities are defined as empirical neighbourhood, colour, and pair measures of a coloured graph. Note that while the *first* component Φ_1 is a continuous function, the *second* component Φ_2 is *discontinuous* in the weak topology because for any b, $\ell(b)$ can be unbounded.

Definition 2.2.3 (Consistency). *We call* $(\varpi, \nu) \in \tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$ *sub-consistent if*

$$\langle \nu(\cdot,\ell)\,,\,\ell(\cdot)\rangle(a,b) \le \varpi(b,a), \quad \forall a,b \in \mathcal{X}, \tag{2.2.1}$$

and consistent if equality holds in (2.2.1).

We note that, if ν is the empirical neighbourhood measure and ϖ the empirical pair measure of a coloured graph, (ϖ, ν) is consistent and both sides in (2.2.1) represent

$$\tfrac{1}{n}\,(1 + \mathbb{1}_{\{a=b\}})\,\sharp\big\{\text{edges between vertices of colours } a \text{ and } b\big\}.$$

Definition 2.2.4 (Q). *For every* $(\varpi, \nu) \in \tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$*, we define a probability measure* $Q = Q[\varpi, \nu]$ *on* $\mathcal{X} \times \mathcal{N}(\mathcal{X})$ *by*

$$Q(a\,,\,\ell) := \nu_1(a) \prod_{b \in \mathcal{X}} \frac{e^{-\left[\varpi(a,b)/\nu_1(a)\right]} \left[\varpi(a,b)/\nu_1(a)\right]^{\ell(b)}}{\ell(b)!}.$$

This is the law of a pair $(a, \ell) \in \mathcal{X} \times \mathcal{N}(\mathcal{X})$ *where* a *is distributed according to* ν_1 *and, given the value of* a*, the random variables* $\ell(b)$ *are independently Poisson distributed with parameter*

$$\tfrac{\varpi(a,b)}{\nu_1(a)}.$$

We observe that, if (ϖ, ν) sub-consistent then $(Q[\varpi, \nu],\, \nu)$ is sub-consistent.

By $\omega \ll \mu$ we mean ω is absolute continuous with respect to μ.

Definition 2.2.5.

$$\mathfrak{M}[\omega, \varpi] := \big\{\nu\colon (\varpi, \nu) \textit{ sub-consistent and } \nu_1 = \omega\big\} \subset \mathcal{M}(\mathcal{X} \times \mathcal{N}(\mathcal{X})).$$

Definition 2.2.6. *We define the relative entropy of* $\varpi \in \tilde{\mathcal{M}}(\mathcal{X} \times \mathcal{X})$ *with respect another finite measure* $\tilde{\varpi} \in \tilde{\mathcal{M}}(\mathcal{X} \times \mathcal{X})$ *by*

$$H(\varpi \,\|\, \tilde{\varpi}) := \sum_{a,b \in \mathcal{X}} \varpi(a,b) \log \tfrac{\varpi(a,b)}{\tilde{\varpi}(a,b)},$$

and ∞ *if* $\varpi \not\ll \tilde{\varpi}$.

2.3 LDP for Empirical Neighbourhood Measure of Near-Critical Coloured Graphs

We have now set the stage to state our principal theorem, the large deviation principle for the empirical pair measure and the empirical neighbourhood measure. To state this principle, we define the function $\mathfrak{H}_C$ by

$$\mathfrak{H}_C(\varpi \,\|\, \omega) := H\big(\varpi \,\|\, C\omega \otimes \omega\big) + \|C\omega \otimes \omega\| - \|\varpi\|\,,$$

where

$$C\omega \otimes \omega(a, b) = C(a, b)\omega(a)\omega(a).$$

Theorem 2.3.1. *Suppose that X is a coloured random graph with colour law μ and connection probabilities $p_n \colon \mathcal{X} \times \mathcal{X} \to [0,1]$ satisfying $np_n(a,b) \to C(a,b)$ for some nonzero function $C \colon \mathcal{X} \times \mathcal{X} \to [0,\infty)$. Then, as $n \to \infty$, the pair $(L^2,\, M)$ satisfies a large deviation principle in $\tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$ with good rate function,*

$$J(\varpi,\nu) = \begin{cases} H(\nu \,\|\, Q) + H(\nu_1 \,\|\, \mu) + \frac{1}{2}\mathfrak{H}_C(\varpi \,\|\, \nu_1) & \text{if } (\varpi,\nu) \text{ sub-consistent,} \\ \infty & \text{otherwise.} \end{cases}$$

2.3.1 The Rate function

The rate function can be interpreted as follows: $J(\varpi\,,\nu)$ represents the cost of obtaining an empirical pair measure ϖ and an empirical neighbourhood measure ν, this cost is divided into three sub-costs:

(i) $H(\nu_1 \,\|\, \mu)$ represents the cost of obtaining the colour frequency ν_1, this cost is non-negative and vanishes iff $\nu_1 = \mu$,

(ii) $\frac{1}{2}\mathfrak{H}_C(\varpi \,\|\, \nu_1)$ represent the cost of obtaining an empirical pair measure ϖ if the colour law is ν_1, again this cost is non-negative and vanishes iff

$$\varpi = C\,\nu_1 \otimes \nu_1,$$

(iii) $H(\nu \,\|\, Q)$ represents the cost of obtaining an empirical neighbourhood measure ν if the colour law is ν_1 and the edge law is ϖ, this cost is non-negative and vanishes iff $\nu = Q$.

Consequently, $J(\varpi,\, \nu)$ is non-negative and vanishes iff $\varpi(a,b) = C\,\nu_1 \otimes \nu_1$ and

$$\nu(a,\ell) = \mu(a) \prod_{b\in\mathcal{X}} e^{-(C(a,b)\mu(b))}\, \frac{(C(a,b)\mu(b))^{\ell(b)}}{\ell(b)!},$$

for all $(a,\ell) \in \mathcal{X} \times \mathcal{N}(\mathcal{X})$. This is the law of a pair (a,ℓ) where a is distributed according to μ and, given the value of a, the random variables $\ell(b)$ are independently Poisson distributed with parameter

$$C(a,b)\mu(b).$$

Remark 2 Our large deviation principle implies individual large deviation principles for the measures L^2 and M by the contraction principle. See [DZ98, Theorem 4.2.1]. Note that, the functional relationship $L^2 = \Phi_2(M)$ may break down in the limit, because Φ_2 is discontinuous in the weak topology. The possibility of this effect is responsible for the weak form of the condition (ϖ, ν) sub-consistent in the rate function.

2.3.2 LDP for the Degree Distribution of Erdős-Rényi graph

As the degree distribution D is a continuous function of M, Theorem 2.3.1 and the contraction principle imply a large deviation principle for D. For a classical Erdős-Rényi graph the rate function takes on a particularly simple form.

Corollary 2.3.2. *Suppose D is the degree sequence of an Erdős-Rényi graph with connection probability $p_n \in [0,1]$ satisfying $np_n \to c \in (0,\infty)$. Then D satisfies a large deviation principle, as $n \to \infty$, in the space $\mathcal{M}(\mathbb{N} \cup \{0\})$ with good rate function*

$$\delta(d) = \begin{cases} \frac{1}{2} x \log\left(\frac{x}{c}\right) - \frac{1}{2}x + \frac{c}{2} + H(d \,\|\, q_x), & \text{if } \langle d \rangle \le c, \\ \frac{1}{2}\langle d \rangle \log\left(\frac{\langle d \rangle}{c}\right) - \frac{1}{2}\langle d \rangle + \frac{c}{2} + H(d \,\|\, q_{\langle d \rangle}), & \text{if } c < \langle d \rangle < \infty, \\ \infty & \text{if } \langle d \rangle = \infty, \end{cases} \tag{2.3.1}$$

where, in the case $\langle d \rangle \le c$, the value $x = x(d)$ is the unique solution of

$$x = ce^{-2\left(1 - \frac{\langle d \rangle}{x}\right)},$$

and where q_λ is a Poisson distribution with parameter λ and

$$\langle d \rangle := \sum_{m=0}^{\infty} m d(m).$$

On probability measures d with mean c the rate simplifies to the relative entropy of d with respect to the Poisson distribution of the same mean. In [BGL01, Theorem 7.1] a large deviation principle for the degree sequence is formulated for this situation, albeit with a rather implicitly defined rate function. Moreover, the proof given there contains a serious gap: The exponential equivalence stated in paper, [BGL01, Lemma 7.2], is not proved and we conjecture it does not hold.

To be more specific about the problem (we recall from [BGL01] that the empirical occupancy process in the random allocation model is denoted by X_t^n and the degree distribution in the Erdős-Rényi graph model is given by V_t^n); the coupling argument carried out to establish the exponential equivalence of the laws of X_t^n and V_t^n only led to a weak bound on the joint probability. i.e. the copies Y_t^n, W_t^n (of X_t^n and V_t^n resp.) constructed only satisfy for any $\varepsilon > 0$

$$\mathbb{P}\Big\{ \sup_{t\in[0,T]} d(Y_{2t}^n, W_t^n) > \varepsilon \Big\} \le e^{-n\alpha},$$

where α is a non-negative constant.

2.4 Large Deviation Principle for the Empirical Colour Measure and Empirical Pair Measure of Sparse Coloured Random Graphs

2.4.1 Overview

Recall that, the empirical colour measure counts the number of vertices of a given colour while the empirical pair measure counts the number of edges or 'interactions' between each pair of colours. We prove joint large deviation principle for these empirical measures of a class of models of sparse coloured random graphs in the weak topology. The major technique used is change of measure. Using the contraction principle, see [DZ98, Theorem 4.2.1], we derive from our result the large deviation principle for the *number of edges per vertex* in the graph and a large deviation property of the Thermodynamics of the Ising model on Erdős-Renyi graph with independent colours. See Example 2.

We structure the section in the following way. Section 2.4.2 contains notation and definitions of some terms used in the subsequent sections. In Section 2.4.3, we state our main principles in the section, Theorem 2.4.4 and Corollaries 2.4.5, 2.4.5.

Section 2.4.4 contains the proof of our first auxiliary principle in the chapter. We begin the section with the statement of the main technique used, change of measure in Subsection 2.4.4. We prove two useful Lemmas 2.4.8 and 2.4.9 in Subsection 2.4.5. The first is based on the *Euler's formula* and the second on *coupling.*

We obtain an upper bound for the first auxiliary principle in a variational formulation. See Subsection 2.4.6. Closely following this, is a solution of the variational problem including the verification of goodness, see Subsection 2.4.7. The last part of the section, Subsection 2.4.4, covers the proof of the lower bound. It is derived from the upper bound with the transformed measure taking the place of the original law of the graph.

We complete the chapter by deriving all corollaries from Theorem 2.4.4 in Subsection 2.4.9.

2.4.2 Some Definitions

By $\mathcal{C}_1$ and $\mathcal{C}_2$ we denote the space of functions on $\mathcal{X}$ and the space of symmetric functions on $\mathcal{X} \times \mathcal{X}$ respectively.

Definition 2.4.1. *Let $g \in \mathcal{C}_2$ and $f \in \mathcal{C}_1$. For $(\omega, \varpi) \in \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X})$,*

$$\langle \varpi,\, g\rangle := \sum_{a,b \in \mathcal{X}} \varpi(a,b) g(a,b) \qquad \text{and} \qquad \langle \omega,\, f\rangle := \sum_{a \in \mathcal{X}} \omega(a) f(a).$$

Definition 2.4.2. *The entropy of a probability vector $\omega \in \mathcal{M}(\mathcal{X})$ is*

$$H(\omega) := -\sum_{a \in \mathcal{X}} \omega(a) \log \omega(a).$$

Definition 2.4.3. *The relative entropy of a probability measure $\omega \in \mathcal{M}(\mathcal{X})$ with respect to $\mu \in \mathcal{M}(\mathcal{X})$ is*

$$H(\omega \,\|\, \mu) := \sum_{a \in \mathcal{X}} \omega(a) \log \tfrac{\omega(a)}{\mu(a)},$$

and ∞ if $\omega \not\ll \mu$.

2.4.3 First Auxiliary Principle

Recall that the empirical pair measure L^2 on $\mathcal{X} \times \mathcal{X}$ in the case when $a_n^{-1} = n$ is given by

$$L^2(a,b) = \frac{1}{n} \sum_{(u,v)\in E} [\delta_{(X(v),X(u))} + \delta_{(X(u),X(v))}](a,b), \qquad \text{for } a,b \in \mathcal{X}.$$

Observe that this situation corresponds to the near-critical case. We state the joint large deviation principle for the *empirical measure* L^1 and the *empirical pair measure* L^2 in the weak topology.

Theorem 2.4.4. *Suppose that X is a coloured random graph with colour law μ and connection probabilities $p_n\colon \mathcal{X} \times \mathcal{X} \to [0,1]$ satisfying $np_n(a,b) \to C(a,b)$ for some function $C\colon \mathcal{X} \times \mathcal{X} \to [0,\infty)$. Then, as $n \to \infty$, the pair (L^1, L^2) satisfies a large deviation principle in $\mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X})$ with speed n and good rate function,*

$$I(\omega, \varpi) = H(\omega \,\|\, \mu) + \tfrac{1}{2}\mathfrak{H}_C(\varpi \,\|\, \omega). \tag{2.4.1}$$

The rate function can be interpreted as follows: The term $H(\omega \,\|\, \mu)$ represents the *cost* of obtaining the colour frequency ω, this cost is non-negative and vanishes iff $\omega = \mu$. The term $\frac{1}{2}\mathfrak{H}_C(\varpi \,\|\, \omega)$ is the cost of obtaining an empirical pair measure ν if the colour law is ω, again this cost is non-negative and vanishes iff $\varpi = C\,\omega \otimes \omega$, as we will show in the course of the proof.

By contraction, see e.g. [DZ98, Theorem 4.2.1], one can obtain a large deviation principle for L^2 and the number of edges per vertex $|E|/n$ alone as follows.

Corollary 2.4.5. *Suppose that X is a coloured random graph with colour law μ and connection probabilities $p_n\colon \mathcal{X} \times \mathcal{X} \to [0,1]$ satisfying $np_n(a,b) \to C(a,b)$ for some function $C\colon \mathcal{X} \times \mathcal{X} \to [0,\infty)$. Then, as $n \to \infty$,*

(i) L^2 satisfies a large deviation principle in $\tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X})$ with speed n and good rate function,

$$I^2(\varpi) = \inf\big\{I(\omega,\varpi)\colon \omega \in \mathcal{M}(\mathcal{X})\big\}. \tag{2.4.2}$$

(ii) the number of edges per vertex $|E|/n$ satisfies a large deviation principle in $[0,\infty)$ with speed n and convex rate function,

$$\zeta(x) = x\log x - x + \inf_{y>0}\big\{\mathfrak{H}(y) - x\log(\tfrac{1}{2}y) + \tfrac{1}{2}y\big\},$$

where $\mathfrak{H}(y) = \inf H(\omega \,\|\, \mu)$ over all probability vectors ω with $\omega^T C\omega = y$.

Example 1 In the Erdős-Renyi case $C(a,a) = c$ one obtains $\mathfrak{H}(y) = 0$ for $y = c$, and $\mathfrak{H}(y) = \infty$ otherwise. Hence

$$\zeta(x) = x \log x - x - x \log(\tfrac{c}{2}) + \tfrac{c}{2},$$

which is the Cramér rate function for the Poisson distribution with parameter $\frac{c}{2}$.

Example 2 We look at the Erdős-Rényi graph with connection probabilities p_n satisfying $np_n \to c \in (0,\infty)$ and study the random partition function for the *Ising model* on the graph, which is defined as

$$Z(\beta) := \sum_{\eta \in \{-1,+1\}^V} \exp\Big(-\beta \sum_{(u,v)\in E} \eta(u)\eta(v)\Big) \text{ for the inverse temperature } \beta > 0.$$

Denoting by E expectation with respect to the graph, we note that

$$\mathsf{E}Z(\beta) = 2^n \mathbb{E} \exp\Big(-n\tfrac{\beta}{2} \int xy\, L^2(dx\,, dy)\Big)\,,$$

where $\mathbb{E}$ is expectation with respect to the graph randomly coloured using independent colours chosen uniformly from $\{-1,1\}$.

Then Varadhan's lemma, see e.g. [DZ98, Theorem 4.3.1] and a simple manipulation give that

$$\begin{aligned}
&\lim_{n\to\infty} \tfrac{1}{n} \log \mathsf{E}Z(\beta) \\
&= \log 2 + \sup\Big\{ \tfrac{\beta}{2} \int xy\, \varpi(dx\,, dy) - I(\omega, \varpi) \,:\, \omega \in \mathcal{M}(\mathcal{X}),\, \varpi \in \mathcal{M}_*(\{-1,1\}^2)\Big\} \\
&= \sup\Big\{ \tfrac{\beta}{2}\,(\varpi(\Delta) - \varpi(\Delta^c)) - x\log(x) - (1-x)\log(1-x) - \tfrac{1}{2}\big(H(\varpi \,\|\, \omega_x) + c\big) \\
&\qquad - \tfrac{1}{2}\,\|\varpi\|\Big\},
\end{aligned}$$

where Δ is the diagonal in $\{-1,1\}^2$, and the supremum is over all $x \in [0,1]$ and $\varpi \in \mathcal{M}_*(\{-1,1\}^2)$, and the measure $\omega_x \in \tilde{\mathcal{M}}_*(\{-1,1\}^2)$ is defined by

$$\omega_x(i,j) = c x^{(2+i+j)/2}(1-x)^{(2-i-j)/2} \qquad \text{for } i,j \in \{-1,1\}\,.$$

Note that the last expression is an optimization problem in only four real variables.

2.4.4 The Technique of Change-of-Measure

Definition 2.4.6. *Given the function $\tilde{f} \in \mathcal{C}_1$ define the constant $U_{\tilde{f}}$ by*

$$U_{\tilde{f}} = \log \sum_{a\in\mathcal{X}} e^{\tilde{f}(a)}\mu(a).$$

Definition 2.4.7. *For $\tilde{g} \in \mathcal{C}_2$ define the function $\tilde{h}_n \colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ by*

$$\tilde{h}_n(a,b) = \log\Big[\big(1 - p_n(a,b) + p_n(a,b)e^{\tilde{g}(a,b)}\big)^{-n}\Big], \qquad \textit{for } a,b \in \mathcal{X}.$$

We use $\tilde{f}$ and $\tilde{g}$ to define (for sufficiently large n) a new random graph as follows:

- For the n labeled vertices V assign colours from $\mathcal{X}$ independently and identically according to the colour law $\widetilde{\mu}$ defined by

$$\widetilde{\mu}(a) = e^{\tilde{f}(a)-U_{\tilde{f}}}\mu(a).$$

- Given any two vertices $u, v \in V$, with u carrying colour a and v carrying colour b connect vertex u to vertex v with probability

$$\tilde{p}_n(a,b) = \frac{p_n(a,b)e^{\tilde{g}(a,b)}}{1 - p_n(a,b) + p_n(a,b)e^{\tilde{g}(a,b)}},$$

otherwise keep u and v disconnected.

Denote the transformed law by $\tilde{\mathbb{P}}$ and define $L^1_\Delta := \dfrac{1}{n}\displaystyle\sum_{u\in V}\delta_{(X(u),X(u))}$.

We recall that $\mathcal{E} = E \cup E^c$, is the set of all possible links. We observe that $\tilde{\mu}$ is a probability measure, $\tilde{p}_n(a,\, b) \in [0,\, 1]$, $\forall\, a, b \in \mathcal{X}$ and that $\tilde{\mathbb{P}}$ is absolutely continuous with respect to $\mathbb{P}$, as for a coloured random graph X,

$$\begin{aligned}
\frac{d\tilde{\mathbb{P}}}{d\mathbb{P}}(X) &= \prod_{u\in V}\tfrac{\tilde{\mu}(X(u))}{\mu(X(u))}\prod_{(u,v)\in E}\tfrac{\tilde{p}_n(X(u),X(v))}{p_n(X(u),X(v))}\prod_{(u,v)\notin E}\tfrac{1-\tilde{p}_n(X(u),X(v))}{1-p_n(X(u),X(v))}\\
&= \prod_{u\in V}\tfrac{\tilde{\mu}(X(u))}{\mu(X(u))}\prod_{(u,v)\in E}\tfrac{\tilde{p}_n(X(u),X(v))}{p_n(X(u),X(v))}\times\tfrac{n-np_n(X(u),X(v))}{n-n\tilde{p}_n(X(u),X(v))}\prod_{(u,v)\in\mathcal{E}}\tfrac{n-n\tilde{p}_n(X(u),X(v))}{n-np_n(X(u),X(v))}\\
&= \prod_{u\in V} e^{\tilde{f}(X(u))-U_{\tilde{f}}}\prod_{(u,v)\in E} e^{\tilde{g}(X(u),X(v))}\prod_{(u,v)\in\mathcal{E}} e^{\frac{1}{n}\tilde{h}_n(X(u),X(v))}\\
&= e^{n\langle L^1,\tilde{f}-U_{\tilde{f}}\rangle+n\langle\frac{1}{2}L^2,\tilde{g}\rangle+n\langle\frac{1}{2}L^1\otimes L^1,\tilde{h}_n\rangle-\langle\frac{1}{2}L^1_\Delta,\tilde{h}_n\rangle}.
\end{aligned} \tag{2.4.3}$$

2.4.5 Some Useful Lemmas

Lemma 2.4.8 (Euler's Formula). *If* $a_n^{-1}p_n(a,b) \to C(a,b)$, $\forall a,b \in \mathcal{X}$ *and* $(a_n) \to 0$, *then*

$$\lim_{n\to\infty}\left[1+\alpha p_n(a,b)\right]^{a_n^{-1}} = e^{\alpha C(a,b)},\ \forall a,b \in \mathcal{X} \text{ and } \alpha \in \mathbb{R}. \tag{2.4.4}$$

Proof. Observe that, for any $\varepsilon > 0$ and for large n we have

$$\Big[1+a_n(\alpha C(a,b)-\varepsilon)\Big]^{a_n^{-1}} \le \Big[1+\alpha p_n(a,b)\Big]^{a_n^{-1}} \le \Big[1+a_n(\alpha C(a,b)+\varepsilon)\Big]^{a_n^{-1}}, \tag{2.4.5}$$

by the point wise convergence. Hence by the sandwich theorem and Euler's formula we have (2.4.4). ■

Lemma 2.4.9 (Exponential Tightness). *For every* $\alpha > 0$, $\exists N \in \mathbb{N}$ *such that*

$$\limsup_{n\to\infty} \tfrac{1}{a_n n^2} \log \mathbb{P}\Big\{|E| > a_n n^2 N\Big\} \le -\alpha. \tag{2.4.6}$$

Proof. Let $c > \max_{a,b\in\mathcal{X}} C(a,b) > 0$. Using Chebysheff's inequality and Lemma 2.4.8, we have (for sufficiently large n)

$$\begin{aligned}
\mathbb{P}\Big\{|E| \ge a_n n^2 l\Big\} &\le e^{-a_n n^2 l}\mathbb{E}\Big\{e^{|E|}\Big\} \\
&\le e^{-a_n n^2 l} \sum_{k=0}^{n(n-1)/2} e^k \binom{n(n-1)/2}{k} \Big(a_n c\Big)^k \Big(1-a_n c\Big)^{n(n-1)/2-k} \\
&= e^{-a_n n^2 l}\Big(1+(e-1)a_n c\Big)^{a_n^{-1}(a_n n(n-1)/2)} \\
&\le e^{-a_n n^2 l} e^{a_n n^2 (c(e-1+o(1)))}.
\end{aligned}$$

Now given α choose $N \in \mathbb{N}$ such that $N > \alpha + c(e-1)$ and observe that, for sufficiently large n,

$$\mathbb{P}\Big\{|E| \ge a_n n^2 N\Big\} \le e^{-a_n n^2 \alpha},$$

which implies the statement. ■

2.4.6 Upper Bound in the First Auxiliary Principle

Definition 2.4.10. *Define the function* $I\colon \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}(\mathcal{X} \times \mathcal{X}) \to [0,\infty]$ *by*

$$\hat{I}(\omega,\varpi) = \sup_{\substack{f\in\mathcal{C}_1\\ g\in\mathcal{C}_2}} \Big\{ \sum_{a\in\mathcal{X}} (f(a) - U_f)\omega(a) + \tfrac{1}{2}\sum_{a,b\in\mathcal{X}} g(a,b)\varpi(a,b) + \tfrac{1}{2}\sum_{a,b\in\mathcal{X}} (1-e^{g(a,b)})C(a,b)\omega(a)\omega(b)\Big\}.$$

Definition 2.4.11. *Given* $\tilde{g} \in \mathcal{C}_2$ *define* $\tilde{\beta}\colon \mathcal{X}\times\mathcal{X} \to \mathbb{R}$ *by*

$$\tilde{\beta}(a,b) = (1-e^{\tilde{g}(a,b)})C(a,b).$$

The upper bound in a variational formulation. To start, we recall that $\mathbb{E}$ is the expectation with respect to $\mathbb{P}$.

Lemma 2.4.12. *For each closed set* $F \subset \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}(\mathcal{X}\times\mathcal{X})$,

$$\limsup_{n\to\infty} \tfrac{1}{n} \log \mathbb{P}\big\{(L^1,L^2) \in F\big\} \le - \inf_{(\omega,\varpi)\in F} \hat{I}(\omega,\varpi).$$

Proof. Fix $\tilde{f} \in \mathcal{C}_1$, $\tilde{g} \in \mathcal{C}_2$ and observe from Lemma 2.4.8 that,

$$\lim_{n\to\infty} \tilde{h}_n(a,b) = \tilde{\beta}(a,b), \ \forall a,b \in \mathcal{X}.$$

Using (2.4.3), we obtain the inequality

$$e^{2\max_{a\in\mathcal{X}}|\beta(a,a)|} \ge \int e^{\langle \frac{1}{2}L^1_{\Delta}, \tilde{h}_n\rangle} d\tilde{\mathbb{P}} = \mathbb{E}\Big\{ e^{n\langle L^1, \tilde{f}-U_{\tilde{f}}\rangle + n\langle \frac{1}{2}L^2, \tilde{g}\rangle + n\langle \frac{1}{2}L^1\otimes L^1, \tilde{h}_n\rangle}\Big\}.$$

Therefore, we have that

$$\limsup_{n\to\infty} \tfrac{1}{n} \log \mathbb{E}\Big\{ e^{n\langle L^1, \tilde{f}-U_{\tilde{f}}\rangle + n\langle \frac{1}{2}L^2, \tilde{g}\rangle + n\langle \frac{1}{2}L^1\otimes L^1, \tilde{h}_n\rangle}\Big\} \le 0. \tag{2.4.7}$$

We fix $\varepsilon > 0$ and define the $\varepsilon-$ rate function by $\hat{I}_\varepsilon(\omega,\varpi) = \min\{\hat{I}(\omega,\varpi), \varepsilon^{-1}\} - \varepsilon$.

Suppose that $(\omega,\varpi) \in F$, choose $\tilde{f} \in \mathcal{C}_1$ and $\tilde{g} \in \mathcal{C}_2$ such that

$$\langle \omega, \, \tilde{f} - U_{\tilde{f}}\rangle + \tfrac{1}{2}\langle \varpi, \, \tilde{g}\rangle + \tfrac{1}{2}\langle \tilde{\beta}, \omega\otimes\omega\rangle \ge \hat{I}_\varepsilon(\omega,\varpi).$$

By finiteness of $\mathcal{X}$, the mapping

$$(\tilde{\omega}, \tilde{\varpi}) \mapsto \langle \tilde{\omega},\, \tilde{f} - U_{\tilde{f}} \rangle + \tfrac{1}{2} \langle \tilde{\varpi},\, \tilde{g} \rangle + \tfrac{1}{2} \langle \tilde{\omega} \otimes \tilde{\omega},\, \tilde{\beta} \rangle$$

is continuous in the weak topology. Therefore, we can find open neighbourhoods B^2_{ϖ} and B^1_{ω} of ϖ, ω such that

$$\inf_{\tilde{\omega} \in B^1_{\omega},\, \tilde{\varpi} \in B^2_{\varpi}} \Big\{ \langle \tilde{\omega},\, \tilde{f} - U_{\tilde{f}} \rangle + \tfrac{1}{2} \langle \tilde{\varpi},\, \tilde{g} \rangle + \tfrac{1}{2} \langle \tilde{\omega} \otimes \tilde{\omega},\, \tilde{\beta} \rangle \Big\} \geq \hat{I}_{\varepsilon}(\omega, \varpi) - \varepsilon.$$

Using Chebysheff's inequality and (2.4.7) we have that

$$\begin{aligned} &\limsup_{n \to \infty} \tfrac{1}{n} \log \mathbb{P}\big\{ (L^1, L^2) \in B^1_{\omega} \times B^2_{\varpi} \big\} \\ &\qquad \leq \limsup_{n \to \infty} \tfrac{1}{n} \log \mathbb{E}\Big\{ e^{n\langle L^1, \tilde{f} - U_{\tilde{f}} \rangle + n\langle \frac{1}{2} L^2, \tilde{g} \rangle + n \langle \frac{1}{2} L^1 \otimes L^1, \tilde{h}_n \rangle} \Big\} - \hat{I}_{\varepsilon}(\omega, \varpi) + \varepsilon \\ &\qquad \leq -\hat{I}_{\varepsilon}(\omega, \varpi) + \varepsilon. \end{aligned} \tag{2.4.8}$$

Now we use Lemma 2.4.9 with $M = \varepsilon^{-1}$, to choose $N(\varepsilon) \in \mathbb{N}$ such that

$$\limsup_{n \to \infty} \tfrac{1}{n} \log \mathbb{P}\big\{ |E| > nN(\varepsilon) \big\} \leq -\varepsilon^{-1}.$$

For this $N(\varepsilon)$, define the set $K_{N(\varepsilon)}$ by

$$K_{N(\varepsilon)} = \big\{ (\omega, \varpi) \in \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}(\mathcal{X} \times \mathcal{X}) : \|\varpi\| \leq 2N(\varepsilon) \big\}.$$

The set $K_{N(\varepsilon)} \cap F$ is compact and therefore may be covered by finitely many sets

$$B^1_{\omega_r} \times B^2_{\varpi_r}, \dots, B^1_{\omega_m} \times B^2_{\varpi_m} \;\text{ with }\; (\omega_r, \varpi_r) \in F \text{ for } \; r = 1, ..., m.$$

Consequently,

$$\begin{aligned} \mathbb{P}\Big\{ (L^1, L^2) \in F \Big\} &\leq \mathbb{P}\Big\{ (L^1, L^2) \in \bigcup_{r=1}^{m} B^1_{\omega_r} \times B^2_{\varpi_r} \Big\} + \mathbb{P}\Big\{ (L^1, L^2) \notin K_{N(\varepsilon)} \Big\} \\ &\leq \sum_{r=1}^{m} \mathbb{P}\Big\{ (L^1, L^2) \in B^1_{\omega_r} \times B^2_{\varpi_r} \Big\} + \mathbb{P}\Big\{ (L^1, L^2) \notin K_{N(\varepsilon)} \Big\}. \end{aligned}$$

We may now use (2.4.8) to obtain

$$\begin{aligned} \limsup_{n \to \infty} \tfrac{1}{n} \log \mathbb{P}\big\{ (L^1, L^2) \in F \big\} &\leq \max_{r=1}^{m} \limsup_{n \to \infty} \tfrac{1}{n} \log \mathbb{P}\big\{ (L^1, L^2) \in B^1_{\omega_r} \times B^2_{\varpi_r} \big\} \vee -\tfrac{1}{\varepsilon} \\ &\leq -\inf_{(\omega, \varpi) \in F} \hat{I}_{\varepsilon}(\omega, \varpi) \vee -\tfrac{1}{\varepsilon} + \varepsilon. \end{aligned}$$

Taking $\varepsilon \downarrow 0$ we have the desired statement. ∎

2.4.7 Identification of the Rate Function

We express the rate function in term of relative entropies, see for example [DZ98, (2.15)], and consequently show that it is a good rate function. To begin, recall that for any $(\omega, \varpi) \in \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}(\mathcal{X} \times \mathcal{X})$, is given by $\mathfrak{H}_C$

$$\mathfrak{H}_C(\varpi \,\|\, \omega) := H\big(\varpi \,\|\, C\omega \otimes \omega\big) + \|C\omega \otimes \omega\| - \|\varpi\| \,.$$

Lemma 2.4.13.

(i) $\hat{I}(\omega, \varpi) = I(\omega, \varpi)$, *for any* $(\omega, \varpi) \in \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}(\mathcal{X} \times \mathcal{X})$,

(ii) I *is a convex, good rate function and*

(iii) $\mathfrak{H}_C(\varpi \,\|\, \omega) \geq 0$ *with equality if and only if* $\varpi = C\omega \otimes \omega$. *In particular, it is a good rate function.*

Proof. (i) Suppose that $\varpi \not\ll C\omega \otimes \omega$. Then, there exists $a_0, b_0 \in \mathcal{X}$ with

$$C\omega \otimes \omega(a_0, b_0) = 0 \qquad \textit{and} \qquad \varpi(a_0, b_0) > 0.$$

Define $\hat{g} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ by

$$\hat{g}(a, b) = \log\big[K(\mathbb{1}_{(a_0,b_0)}(a, b) + \mathbb{1}_{(b_0,a_0)}(a, b)) + 1\big],$$

for $a, b \in \mathcal{X}$ and $K > 0$.

For this choice of $\hat{g}$ and $f = 0$ we have

$$\sum_{a\in\mathcal{X}} \big(f(a) - U_f\big)\omega(a) + \sum_{a,b\in\mathcal{X}} \tfrac{1}{2}\hat{g}(a, b)\varpi(a, b) + \sum_{a,b\in\mathcal{X}} \tfrac{1}{2}(1 - e^{\hat{g}(a,b)})C(a, b)\omega(a)\omega(b)$$
$$= \tfrac{1}{2}\log(K + 1)\varpi(a_0, b_0) \xrightarrow{K\uparrow\infty} \infty.$$

Now suppose that $\varpi \ll C\omega \otimes \omega$. Then, we have

$$\hat{I}(\omega, \varpi) = \sup_{f\in\mathcal{C}_1} \Big\{ \sum_{a\in\mathcal{X}} \Big(f(a) - \log \sum_{a\in\mathcal{X}} e^{f(a)}\mu(a)\Big)\,\omega(a)\Big\} + \tfrac{1}{2} \sum_{a,b\in\mathcal{X}} C(a, b)\omega(a)\omega(b)$$
$$+ \tfrac{1}{2} \sup_{g\in\mathcal{C}_2} \Big\{ \sum_{a,b\in\mathcal{X}} g(a, b)\varpi(a, b) - \sum_{a,b\in\mathcal{X}} e^{g(a,b)}C(a, b)\omega(a)\omega(b)\Big\}.$$

By the variational characterization of entropy, the first term equals $H(\omega \,\|\, \mu)$. By the substitution $h = e^g \, \frac{C\omega\otimes\omega}{\varpi}$ we obtain

$$\begin{aligned}
&\sup_{g\in\mathcal{C}_2}\Big\{\sum_{a,b\in\mathcal{X}} g(a,b)\varpi(a,b) - \sum_{a,b\in\mathcal{X}} e^{g(a,b)}C(a,b)\omega(a)\omega(b)\Big\}.\\
&\sup_{\substack{h\in\mathcal{C}_2\\ h\geq 0}}\sum_{a,b\in\mathcal{X}}\Big[\log\Big(h(a,b)\frac{\varpi(a,b)}{C(a,b)\omega(a)\omega(b)}\Big) - h(a,b)\Big]\varpi(a,b)\\
&= \sup_{\substack{h\in\mathcal{C}_2\\ h\geq 0}}\sum_{a,b\in\mathcal{X}}\big(\log h(a,b) - h(a,b)\big)\varpi(a,b) + \sum_{a,b\in\mathcal{X}}\log\Big(\frac{\varpi(a,b)}{C(a,b)\omega(a)\omega(b)}\Big)\varpi(a,b)\\
&= -\|\varpi\| + H(\varpi \,\|\, C\omega\otimes\omega),
\end{aligned}$$

where we have used $\sup_{x>0}\log x - x = -1$ in the last step. This yields

$$\hat{I}(\omega,\varpi) = I(\omega,\varpi).$$

(ii) Recall from (2.4.1) and the definition of $\mathfrak{H}_C$ that

$$I(\omega,\varpi) = H(\omega\,\|\,\mu) + \tfrac{1}{2}H\big(\varpi\,\|\,C\omega\otimes\omega\big) + \tfrac{1}{2}\|C\omega\otimes\omega\| - \tfrac{1}{2}\|\varpi\|.$$

All summands are continuous in ω,ϖ and thus I is a rate function. Moreover, for all $\alpha<\infty$, the level sets $\{I\leq\alpha\}$ are contained in the bounded set

$$\{(\omega,\varpi)\in\mathcal{M}(\mathcal{X})\times\tilde{\mathcal{M}}(\mathcal{X}\times\mathcal{X})\colon\ \mathfrak{H}_C(\varpi\,\|\,\omega)\leq\alpha\}$$

and are therefore compact. Consequently, I is a good rate function.
(iii) Consider the non-negative function $\xi(x) = x\log x - x + 1$, for $x>0$, $\xi(0)=1$, which has its only root in $x=1$. Note that

$$\mathfrak{H}_C(\varpi\,\|\,\omega) = \begin{cases}\int \xi\circ g \; dC\omega\otimes\omega & \text{if } g := \frac{d\varpi}{dC\omega\otimes\omega}\geq 0 \text{ exists}\\ \infty & \text{otherwise.}\end{cases} \tag{2.4.9}$$

Hence $\mathfrak{H}_C(\varpi\,\|\,\omega)\geq 0$, and, if $\varpi = C\omega\otimes\omega$, then $\xi(\frac{d\varpi}{dC\omega\otimes\omega}) = \xi(1) = 0$ and so $\mathfrak{H}_C(C\omega\otimes\omega\,\|\,\omega) = 0$. Conversely, if $\mathfrak{H}_C(\varpi\,\|\,\omega) = 0$, then $\varpi(a,b)>0$ implies $C\omega\otimes\omega(a,b)>0$, which then implies $\xi\circ g(a,b) = 0$ and further $g(a,b) = 1$. Hence $\varpi = C\omega\otimes\omega$, which completes the proof of (iii). ∎

2.4.8 Lower Bound in the First Auxiliary Principle

Lemma 2.4.14. *For every open set $O \subset \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}(\mathcal{X} \times \mathcal{X})$,*

$$\liminf_{n\to\infty} \tfrac{1}{n} \log \mathbb{P}\big\{(L^1, L^2) \in O\big\} \geq - \inf_{(\omega,\varpi)\in O} I(\omega, \varpi).$$

Proof. Suppose $(\omega, \varpi) \in O$, with $\varpi \ll C\omega \otimes \omega$. Define $\tilde{f}_\omega \colon \mathcal{X} \to \mathbb{R}$ by

$$\tilde{f}_\omega(a) = \begin{cases} \log \frac{\omega(a)}{\mu(a)}, & \text{if } \omega(a) > 0, \\ 0, & \text{otherwise.} \end{cases}$$

and $\tilde{g}_\varpi \colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ by

$$\tilde{g}_\varpi(a,b) = \begin{cases} \log \frac{\varpi(a,b)}{C(a,b)\omega(a)\omega(b)}, & \text{if } \varpi(a,b) > 0, \\ 0, & \text{otherwise.} \end{cases}$$

In addition, we let $\tilde{\beta}_\varpi(a,b) = C(a,b)(1 - e^{\tilde{g}_\varpi(a,b)})$ and note that

$$\tilde{\beta}_\varpi(a,b) = \lim_{n\to\infty} \tilde{h}_{\varpi,n}(a,b) = \lim_{n\to\infty} \log\Big[\big(1 - p_n(a,b) + p_n(a,b) e^{\tilde{g}_\varpi(a,b)}\big)^{-n}\Big], \ \ \forall a, b \in \mathcal{X}.$$

Choose B^1_ω, B^2_ϖ open neighbourhoods of ω, ϖ, such that $B^1_\omega \times B^2_\varpi \subset O$ and $\forall (\tilde{\omega}, \tilde{\varpi}) \in B^1_\omega \times B^2_\varpi$,

$$\langle \omega,\, \tilde{f}_\omega\rangle + \tfrac{1}{2}\langle \varpi,\, \tilde{g}_\varpi\rangle + \tfrac{1}{2}\langle \omega\otimes\omega,\, \tilde{\beta}_\varpi\rangle - \varepsilon \leq \langle \tilde{\omega},\, \tilde{f}_\omega\rangle + \tfrac{1}{2}\langle \tilde{\varpi},\, \tilde{g}_\varpi\rangle + \tfrac{1}{2}\langle \tilde{\omega}\otimes\tilde{\omega},\, \tilde{\beta}_\varpi\rangle.$$

We now use $\tilde{\mathbb{P}}$, the probability measure obtained by transforming $\mathbb{P}$ using the functions $\tilde{f}_\omega$, $\tilde{g}_\varpi$. Note that the colour law in the transformed measure is now ω, and the edge probabilities $\tilde{p}_n(a,b)$ satisfy $n\,\tilde{p}_n(a,b) \xrightarrow{n\uparrow\infty} \frac{\varpi(a,b)}{\omega(a)\omega(b)} =: \tilde{C}(a,b)$. Using (2.4.3), we obtain

$$\begin{aligned}
\mathbb{P}\Big\{(L^1, L^2) \in O\Big\} &\geq \tilde{\mathbb{E}}\Big\{\tfrac{d\mathbb{P}}{d\tilde{\mathbb{P}}}(X)\mathbb{1}_{\{(L^1,L^2)\in B^1_\omega\times B^2_\varpi\}}\Big\} \\
&= \tilde{\mathbb{E}}\Big\{e^{-n\langle L^1,\tilde{f}_\omega\rangle - n\,\langle\frac{1}{2}L^2,\tilde{g}_\varpi\rangle - n\,\langle\frac{1}{2}L^1\otimes L^1,\tilde{g}_\varpi\rangle + \frac{1}{2}\,\langle L^1_\Delta,\tilde{h}_{\varpi,n}\rangle} \times \mathbb{1}_{\{(L^1,L^2)\in B^1_\omega\times B^2_\varpi\}}\Big\} \\
&\geq e^{-n\langle\omega,\tilde{f}_\omega\rangle - n\langle\frac{1}{2}\varpi,\tilde{g}_\varpi\rangle - n\langle\frac{1}{2}\omega\otimes\omega,\tilde{\beta}_\varpi\rangle + m - n\varepsilon} \times \tilde{\mathbb{P}}\Big\{(L^1, L^2) \in B^1_\omega \times B^2_\varpi\Big\},
\end{aligned}$$

where $m = 0 \wedge \min_{a\in\mathcal{X}} \tilde{\beta}_\varpi(a,a)$. Therefore, we have

$$\begin{aligned}
\liminf_{n\to\infty} \tfrac{1}{n} \log \mathbb{P}\big\{(L^1, L^2) \in O\big\} \geq &-\langle \tilde{f}_\omega, \omega\rangle - \tfrac{1}{2}\langle \tilde{g}_\varpi, \varpi\rangle - \tfrac{1}{2}\langle \tilde{\beta}_\varpi, \omega\otimes\omega\rangle - \varepsilon \\
&+ \liminf_{n\to\infty} \tfrac{1}{n} \log \tilde{\mathbb{P}}\big\{(L^1, L^2) \in B^1_\omega \times B^2_\varpi\big\}.
\end{aligned}$$

The result follows once we prove that

$$\liminf_{n\to\infty} \tfrac{1}{n} \log \tilde{\mathbb{P}}\big\{(L^1, L^2) \in B^1_\omega \times B^2_\varpi\big\} = 0. \tag{2.4.10}$$

We use the upper bound (but now with the law $\mathbb{P}$ replaced by $\tilde{\mathbb{P}}$) to prove (2.4.10). Then we obtain

$$\limsup_{n\to\infty} \tfrac{1}{n} \log \tilde{\mathbb{P}}\big\{(L^1, L^2) \in (B^1_\omega \times B^2_\varpi)^c\big\} \le - \inf_{(\tilde{\omega},\tilde{\varpi})\in \tilde{F}} \tilde{I}(\tilde{\omega}, \tilde{\nu}),$$

$\tilde{F} = (B^1_\omega \times B^2_\varpi)^c$ and $\tilde{I}(\tilde{\omega}, \tilde{\varpi}) := H(\tilde{\omega} \,\|\, \omega) + \frac{1}{2}\mathfrak{H}_{\tilde{C}}(\tilde{\varpi} \,\|\, \tilde{\omega})$. Therefore, the proof of the lemma is complete pending the proof of the following Lemma. ■

Lemma 2.4.15. $\inf_{(\tilde{\omega},\tilde{\varpi})\in\tilde{F}} \tilde{I}(\tilde{\omega}, \tilde{\nu}) > 0.$

Proof.
Suppose for contradiction that there exists a sequence $(\tilde{\omega}_n, \tilde{\varpi}) \in \tilde{F}$ with $\tilde{I}(\tilde{\omega}_n, \varpi_n) \downarrow 0$. Then, because $\tilde{I}$ is a good rate function and its level sets are compact, and by lower semicontinuity of the mapping $(\tilde{\omega}, \tilde{\varpi}) \mapsto \tilde{I}(\tilde{\omega}, \tilde{\varpi})$ in both ϖ and $\tilde{\omega}$, we can construct a limit point $(\tilde{\omega}, \varpi) \in \tilde{F}$ with $\tilde{I}(\tilde{\omega}, \tilde{\varpi}) = 0$. By Lemma 2.4.13 this implies $H(\tilde{\omega} \,\|\, \omega) = 0$ and $\mathfrak{H}_C(\tilde{\varpi} \,\|\, \tilde{\omega}) = 0$, hence $\tilde{\omega} = \omega$, and $\tilde{\varpi} = \tilde{C}\tilde{\omega} \otimes \tilde{\omega} = \varpi$ contradicting $(\tilde{\omega}, \tilde{\varpi}) \in \tilde{F}$. ■

2.4.9 Derivation of the Corollaries

Proof of Corollary 2.4.5(i). The proof follow from Theorem 2.4.4 by applying the contraction principle to the continuous linear map $G : \mathcal{M}(\mathcal{X}) \times \mathcal{M}_*(\mathcal{X} \times \mathcal{X}) \to \mathcal{M}_*(\mathcal{X} \times \mathcal{X})$ given by

$$G(\omega, \varpi) = \varpi.$$

Theorem 2.4.4 implies large deviation principle for $G(L^1,\, L^2) = L^2$ with good rate function

$$I^2(\varpi) = \inf\{I(\omega, \varpi) : \, G(\omega, \varpi) = \varpi\}.$$

Proof of Corollary 2.4.5(ii). We begin the proof by defining the continuous linear map $W\colon \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}(\mathcal{X} \times \mathcal{X}) \to [0, \infty)$ by

$$W(\omega, \tilde{\varpi}) = \tfrac{1}{2}\|\tilde{\varpi}\|,$$

where $\|\tilde{\varpi}\|$ is the total mass of the measure $\tilde{\varpi}$. We infer from Theorem 2.4.5 and the contraction principle that $W(L^1, L^2) = |E|/n$, satisfies a large deviation principle in $[0,\infty)$ with the good rate function

$$\zeta(x) = \inf\left\{I(\omega,\tilde{\varpi})\colon W(\omega,\tilde{\varpi}) = x\right\}.$$

To obtain the form of the rate in the corollary , the preceding infimum is reformulated as unconstraint optimization problem.

$$\inf_{\substack{\varpi\in\mathcal{M}(\mathcal{X}\times\mathcal{X})\\ \omega\in\mathcal{M}(\mathcal{X})}}\left\{H(\omega\,\|\,\mu) + xH(\varpi\,\|\,C\omega\otimes\omega) + x\log 2x + \tfrac{1}{2}\,\|C\omega\otimes\omega\| - x\right\}. \quad (2.4.11)$$

By Jensen's inequality $H(\varpi\,\|\,C\omega\otimes\omega) \geq -\log\|C\omega\otimes\omega\|$, with equality if

$$\varpi = \tfrac{C\omega\otimes\omega}{\|C\omega\otimes\omega\|}.$$

Hence, by symmetry of C we have

$$\begin{aligned}\min_{\varpi\in\mathcal{M}(\mathcal{X}\times\mathcal{X})}&\left\{H(\omega\,\|\,\mu) + xH(\varpi\,\|\,C\omega\otimes\omega) + x\log 2x + \tfrac{1}{2}\,\|C\omega\otimes\omega\| - x\right\}\\ &= H(\omega\,\|\,\mu) - x\log\|C\omega\otimes\omega\| + x\log 2x + \tfrac{1}{2}\,\|C\omega\otimes\omega\| - x.\end{aligned}$$

The form given in Corollary 2.4.5 follows by defining

$$y = \|C\omega\otimes\omega\|.$$

2.5 LDP for the Empirical Neighbourhood Measure Conditioned to have a given Empirical Colour Measure and Empirical Pair Measure

2.5.1 Second Auxiliary Principle

We state our second intermediate result, the LDP for the empirical neighbourhood measure conditioned to have a given empirical colour measure and empirical pair measure. We recall that

$$\mathfrak{M}[\omega, \varpi] := \big\{\nu\colon (\varpi, \nu) \text{ sub-consistent and } \nu_1 = \omega\big\}.$$

Theorem 2.5.1. *Suppose the sequence $(\omega_n, \varpi_n) \in \mathcal{M}_n(\mathcal{X}) \times \tilde{\mathcal{M}}_n(\mathcal{X} \times \mathcal{X})$ converges to a limit $(\omega, \varpi) \in \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X})$. Let X be a coloured random graph with n vertices conditioned on the event $\{\Phi(M) = (\omega_n, \varpi_n)\}$. Then, as $n \to \infty$, the empirical neighbourhood measure M of X satisfies a large deviation principle in the space $\mathcal{M}(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$ with good rate function*

$$\tilde{J}_{(\omega,\varpi)}(\nu) = \begin{cases} H(\nu \,\|\, Q) & \text{if } \nu \in \mathfrak{M}[\omega, \varpi] \\ \infty & \text{otherwise.} \end{cases} \tag{2.5.1}$$

Example, in the Erdős-Rényi graph model on n, where cn edges are inserted at random among n vertices $M = D$ and $\Phi(D) = nc$. Therefore, the degree distribution obeys the following LDP in the space of probability measures on $\mathbb{N}\times\{0\}$. To state this principle, we denote by $\mathcal{G}(n, nc)$ the space of all these graphs.

Corollary 2.5.2. *Suppose X is a random graph chosen uniformly from $\mathcal{G}(n, nc)$. Then, as $n \to \infty$, D obeys an LDP in $\mathcal{M}(\mathbb{N} \times \{0\})$ with good rate function*

$$\hat{\xi}(d) = \begin{cases} H(d \,\|\, q_c), & \text{if } \langle d \rangle \le c, \\ \infty & \text{otherwise.} \end{cases}$$

where q_λ is a poisson distribution with mean λ.

We note that on the set of probability measures with mean less or equal c the rate functions in Lemma 2.3.2 and Lemma 2.5.2 concise.

2.5.2 The New Model for the Method of Types

Throughout the proof we may assume that $\omega(a) > 0$ for all $a \in \mathcal{X}$. It is easy to see that the law of the coloured random graph conditioned to have empirical colour measure ω_n and empirical pair measure ϖ_n,

$$\mathbb{P}_{(\omega_n,\varpi_n)} = \mathbb{P}\{\,\cdot \mid \Phi(M) = (\omega_n, \varpi_n)\},$$

can be described in the following manner:

- Assign colours to the vertices by sampling without replacement from the collection of n colours, which contains any colour $a \in \mathcal{X}$ exactly $n\omega_n(a)$ times;
- for every unordered pair $\{a, b\}$ of colours create exactly $n(a, b)$ edges by sampling without replacement from the pool of possible edges connecting vertices of colour a and b, where

$$n(a,b) := \begin{cases} n\,\varpi_n(a,b) & \text{if } a \neq b \\ \frac{n}{2}\,\varpi_n(a,b) & \text{if } a = b\,. \end{cases}$$

For our proof it is convenient to introduce a numbering system, which specifies, for each $\{a, b\}$, the order in which edges are drawn in the second step. More precisely, the edge-number k is attached to both vertices connecting the k^{th} edge. Note that the total number of edge-numbers attached to every vertex corresponds to the degree of the vertex in the graph. All permitted numberings are equally probable, with the total number of possible numberings (given the coloured graph) being

$$\prod_{\{a,b\}} n(a,b)!\;.$$

Notation: Denote by $V(a)$ the set of vertices with colour a, and let $Y_j^{\{a,b\}}$ be the j^{th} edge drawn in the process of connecting vertices of colours $\{a, b\}$. Let $\mathcal{A}_n(\omega_n, \varpi_n)$ be the set of all possible configurations

$$\tilde{X} = \Big(\big(V(a) \,:\, a \in \mathcal{X}\big);\; \big(Y_k^{\{a,b\}} \,:\, k = 1, \ldots, n(a,b);\; \{a,b\} \subset \mathcal{X}\big) \Big)\,.$$

Denote by $\mathcal{B}_n(\omega_n, \varpi_n)$ the set of all coloured graphs x with $L^1(x) = \omega_n$ and $L^2(x) = \varpi_n$.

By $\Psi\colon \mathcal{A}_n(\omega_n, \varpi_n) \to \mathcal{B}_n(\omega_n, \varpi_n)$, we denote canonical mapping which associates the coloured graph to any configuration, i.e. 'forgets' the numbering of the edges.

We write $\mathcal{K}^{(n)}(\omega_n, \varpi_n) := \{M(x) \text{ for some } x \in \mathcal{B}_n(\omega_n, \varpi_n)\}$ for the set of all empirical neighbourhood measures $M(x)$ arising from coloured graphs x with n vertices with

$$\Phi(M(x)) = (\omega_n, \varpi_n).$$

The next inequality is a refinement of Stirling's formula, see [Fel67, Page 52].

Lemma 2.5.3. *For any $n \in \mathbb{N}$, $n^n e^{-n} \leq n! \leq (2\pi n)^{\frac{1}{2}} n^n e^{-n+1/(12n)}$.*

We write $\binom{n}{r_1, \ldots, r_m} := \frac{n!}{r_1! \ldots r_m!}$, where $\sum_{j=1}^m r_j = n$.

2.5.3 A Bound on the Number of Empirical Neighbourhood Measures

In this section we provide an upper bound on the number of measures in $\mathcal{K}^{(n)}(\omega_n, \varpi_n)$. We write m for the number of elements in $\mathcal{X}$. We recall that any $\omega_n \in \mathcal{M}_n(\mathcal{X})$ is of the form

$$\omega_n(a) = \frac{1}{n} \sum_{j=1}^n \delta_{a_j}(a), \qquad \text{for } a_i, a \in \mathcal{X}.$$

Lemma 2.5.4. *There exists $\vartheta = \vartheta(m) > 0$ such that, if $\omega_n \in \mathcal{M}_n(\mathcal{X})$ and $\varpi_n \in \tilde{\mathcal{M}}_n(\mathcal{X} \times \mathcal{X})$, then*

$$\sharp \mathcal{K}^{(n)}(\omega_n, \varpi_n) \leq \exp\left[\vartheta \times (\log n)\,(n\|\varpi_n\|)^{\frac{2m-1}{2m}}\right]. \tag{2.5.2}$$

Let $\mathfrak{I}_m = (\mathbb{N} \cup \{0\})^m$ be the collection of (non-negative) integer vectors of length m. For any $\ell \in \mathfrak{I}_m$ we denote by $\|\ell\|$ its magnitude, i.e. the sum of its entries.

The proof is based on counting integer partitions of vectors. We introduce an ordering $\succcurlyeq$ on $\mathfrak{I}_m$ such that, for any vectors

$$\ell_1 = (\ell_1^{(1)}, \ldots, \ell_1^{(m)}) \text{ and } \ell_2 = (\ell_2^{(1)}, \ldots, \ell_2^{(m)}),$$

we write $\ell_1 \succcurlyeq \ell_2$ if either

(i) $\|\ell_1\| > \|\ell_2\|$, *or*

(ii) $\|\ell_1\| = \|\ell_2\|$ and there is $j \in \{1, \ldots, m\}$ with $\ell_1^{(k)} = \ell_2^{(k)}$, for all $k < j$, and $\ell_1^{(j)} > \ell_2^{(j)}$, *or*

(iii) $\ell_1 = \ell_2$.

A collection $(\ell_1, \ldots, \ell_k)$ of elements in $\mathfrak{I}_m$ is an *integer partition* of the vector $\ell \in \mathfrak{I}_m$, if

$$\ell_1 \succcurlyeq \ldots \succcurlyeq \ell_k \neq 0 \qquad \text{and} \qquad \ell_1 + \ldots + \ell_k = \ell.$$

Any integer partition of a vector $\ell \in \mathfrak{I}_m$ induces an integer partition $\|\ell_1\|, \ldots, \|\ell_k\|$ of its magnitude $\|\ell\|$, which we call its *sum-partition.* We denote by $\mathcal{P}_m(\ell)$ the set of integer partitions of ℓ.

Lemma 2.5.5. *There exists $\vartheta = \vartheta(m) > 0$ such that, for any $\ell \in \mathfrak{I}_m$ of magnitude n,*

$$\sharp\mathcal{P}_m(\ell) \leq \exp\left[\vartheta\,(\log n)\, n^{\frac{2m-1}{2m}}\right].$$

Proof. Let $\ell \in \mathfrak{I}_m$ be a vector of magnitude n and $(\ell_1, \ldots, \ell_k)$ be an integer partition of ℓ. We rewrite the partition as $(\mathfrak{m}_{1,1}, \ldots, \mathfrak{m}_{1,k_1}; \mathfrak{m}_{2,1}, \ldots, \mathfrak{m}_{2,k_2}; \ldots; \mathfrak{m}_{r,1}, \ldots, \mathfrak{m}_{r,k_r})$ such that all vectors in the same block (indicated by the first subscript) have the same magnitude, which we denote $y_1, \ldots, y_r$, and such that $y_1 > \cdots > y_r > 0$. Note that for the block sizes we have $k_1 + \cdots + k_r = k$ and that $k_1 y_1 + \cdots + k_r y_r = n$.

For a moment, look at a fixed block $\mathfrak{m}_{j,1}, \ldots, \mathfrak{m}_{j,k_j}$. It is easy to see that the number of integer vectors of length m and magnitude y_j is given by

$$b(y_j, m) := \binom{y_j + m - 1}{m - 1} \leq c(m)\, y_j^{m-1}.$$

Writing $\mathfrak{m}_{j,0}$ for the largest and $\mathfrak{m}_{j,k_j+1}$ for the smallest of these vectors in the ordering $\succcurlyeq$, we note that

$$p\colon \{0, \ldots, k_j + 1\} \to \{\mathfrak{m} \in \mathfrak{I}_m \colon \|\mathfrak{m}\| = y_j\}, \qquad p(i) = \mathfrak{m}_{j,i},$$

is a non-increasing path of length $k_j + 2$ into an ordered set of size $b(y_j, m)$, which connects the smallest to the largest element. The number of such paths is easily seen to be

$$\binom{b(y_j, m) + k_j}{k_j}.$$

Therefore, the number of integer partitions of ℓ with given sum-partition $(y_1, \overset{k_1}{\ldots}, y_1, \ldots, y_r, \overset{k_r}{\ldots}, y_r)$ is

$$\prod_{j=1}^{r} \binom{b(y_j, m) + k_j}{k_j} \leq \max_{\substack{a_1, \ldots, a_r > 0 \\ \sum a_j = n}} \prod_{j=1}^{r} \left\{ \max_{\substack{y, k \in \mathbb{N} \\ yk = a_j}} \binom{c(m) y^{m-1} + k}{k} \right\}.$$

To maximize the binomial coefficient over the set $yk = a_j$, we distinguish between the cases when (i) $a_j \le c(m)y^m$, (ii) $a_j > c(m)y^m$ and observe that

$$\binom{c(m)y^{m-1} + \frac{a_j}{y}}{\frac{a_j}{y}} \le \begin{cases} \binom{2c(m)y^{m-1}}{\frac{a_j}{y}} & \text{if } a_j \le c(m)y^m, \\ \binom{2\frac{a_j}{y}}{c(m)y^{m-1}} & \text{if } a_j > c(m)y^m. \end{cases}$$

Using the upper bound $\binom{i}{r} \le \left(\frac{ie}{r}\right)^r$, for $r, i \in \mathbb{N}$ with $r \le i$ and the inequality $(\frac{a_j}{c(m)})^{1/m} \le y \le a_j \le n$ we obtain, for some constants $C_0 = C_0(m)$, $C_1 = C_1(m)$,

$$\begin{aligned} \binom{2c(m)y^{m-1}}{a_j/y} \le \left(\frac{2c(m)y^{m-1}e}{a_j/y}\right)^{a_j/y} &\le \exp\left(C_0\,(a_j/y)\,\log n\right) \\ &\le \exp\left((\log n)\,C_1\,a_j^{\frac{m-1}{m}}\right). \end{aligned}$$

The same upper bound for binomial coefficients and $1 \le y \le (\frac{a_j}{c})^{1/m} \le a_j \le n$ yield for some constant $C_2 = C_2(m)$,

$$\binom{2a_j/y}{c(m)y^{m-1}} \le \left(\frac{2(a_j/y)e}{c(m)y^{m-1}}\right)^{c(m)y^{m-1}} \le \exp\left((\log n)\,C_2\,a_j^{\frac{m-1}{m}}\right).$$

From this, we have for some $C = C(m) > 0$, the upper bound

$$\prod_{j=1}^{r} \binom{b(y_j, m) + k_j}{k_j} \le \max_{\substack{a_1,\dots,a_r>0 \\ \sum a_j = n}} \prod_{j=1}^{r} \exp((\log n) C\, a_j^{\frac{m-1}{m}}),$$

which is estimated further (using Hölder's inequality) by

$$\exp\left((\log n)\,C \sum_{j=1}^{r} a_j^{\frac{m-1}{m}}\right) \le \exp\left((\log n)\,C\,r^{\frac{1}{m}} \left(\sum_{j=1}^{r} a_j\right)^{\frac{m-1}{m}}\right).$$

We observe that all y_j are different, positive and that their sum is not greater than n, so we have that $r^2/2 \le 1 + \ldots + r \le y_1 + \ldots + y_r \le n$. Recalling that $a_1 + \cdots + a_r = n$, our upper bound becomes

$$\exp\left((\log n)\,C\,(2n)^{\frac{1}{2m}}\,n^{\frac{m-1}{m}}\right) = \exp\left(\tfrac{\vartheta}{2}\,(\log n)\,n^{\frac{2m-1}{2m}}\right),$$

for some $\vartheta = \vartheta(m) > 0$. Note that from our argument so far one can easily recover the (well-known) fact that the number of integer partitions of n is bounded by $e^{(\vartheta/2)\sqrt{n}}$ first obtained by G. H. Hardy and Ramanujan in 1918.

Combining this with the upper bound for the number of integer partitions with a given sum-partition, we obtain the claim.

■

Proof of Lemma 2.5.4. Suppose $\omega_n \in \mathcal{M}_n(\mathcal{X})$ and $\varpi_n \in \tilde{\mathcal{M}}_n(\mathcal{X} \times \mathcal{X})$. For $a \in \mathcal{X}$, we look at the mappings

$$\Phi_a \colon \mathcal{K}^{(n)}(\omega_n, \varpi_n) \ni \frac{1}{n} \sum_{v \in V} \delta_{(X(v),L(v))} \mapsto \big(L_1^a, \ldots, L_{n\omega(a)}^a\big),$$

where $(L_1^a, \ldots, L_{n\omega(a)}^a)$ is the ordering of the vectors $L(v)$, for all $v \in V$ with $X(v) = a$, and thus constitutes and integer partition of the vector

$$\big(n\varpi_n(a,b) \colon b \in \mathcal{X}\big),$$

which has magnitude $n \sum_b \varpi_n(a,b)$. The combined mapping $\Phi = (\Phi_a \colon a \in \mathcal{X})$ is injective, and therefore, by Lemma 2.5.5,

$$\begin{aligned} \sharp\mathcal{K}^{(n)}(\omega_n, \varpi_n) &\leq \exp\Big[\vartheta \sum_{a \in \mathcal{X}} \log\big(n \sum_{b \in \mathcal{X}} \varpi_n(a,b)\big) \big(n \sum_{b \in \mathcal{X}} \varpi_n(a,b)\big)^{\frac{2m-1}{2m}}\Big] \\ &\leq \exp\Big[\vartheta \log\big(n\|\varpi\|\big) \big(n\|\varpi_n\|\big)^{\frac{2m-1}{2m}}\Big], \end{aligned}$$

where we have used $\varpi_n(a,b) \leq \|\varpi\|$ and $\varpi_n(a,b) \leq \|\varpi_n\|$ in the last step. ■

2.5.4 Upper Bound in the Second Auxiliary Principle

We are now ready to prove an upper bound for the large deviation probability in Theorem 2.5.1, starting with a lemma based on the method of types.

Lemma 2.5.6. *For any sequence (ν_n) with $\nu_n \in \mathcal{K}^{(n)}(\omega_n, \varpi_n)$ we have*

$$\mathbb{P}\big\{M = \nu_n \,\big|\, \Phi(M) = (\omega_n, \varpi_n)\big\} \leq \exp\big(-nH(\nu_n \,\|\, Q_n) + \varepsilon_1^{(n)}(\nu_n)\big),$$

where

$$Q_n(a\,,\,\ell) = \omega_n(a) \prod_{b \in \mathcal{X}} \frac{e^{-\varpi_n(a,b)/\omega_n(a)}[\varpi_n(a,b)/\omega_n(a)]^{\ell(b)}}{\ell(b)!}, \quad \textit{for } \ell \in \mathcal{N}(\mathcal{X}),$$

and

$$\lim_{n\uparrow\infty} \tfrac{1}{n} \sup_{\nu_n \in \mathcal{K}^{(n)}(\omega_n,\varpi_n)} \varepsilon_1^{(n)}(\nu_n) = 0.$$

Proof. The proof of this lemma is based on the method of types, see [DZ98, Chapter 2]. For any $\nu_n \in \mathcal{K}^{(n)}(\omega_n, \varpi_n)$ we have

$$\mathbb{P}\{M = \nu_n \,|\, \Phi(M) = (\omega_n, \varpi_n)\} = \frac{\sharp\{\tilde{x} \in \mathcal{A}_n(\omega_n, \varpi_n) \,:\, M \circ \Psi(\tilde{x}) = \nu_n\}}{\sharp\{\tilde{x} \in \mathcal{A}_n(\omega_n, \varpi_n))\}}. \tag{2.5.3}$$

Now, by elementary counting, the denominator on the right side of (2.5.3) is

$$\binom{n}{n\omega_n(a),\, a \in \mathcal{X}} \prod_{\{a,b\}} \prod_{k=1}^{n(a,b)} \Big(\frac{n^2\omega_n(a)\omega_n(b) - n\omega_n(a)\mathbb{1}_{\{a=b\}}}{1 + \mathbb{1}_{\{a=b\}}} - (k-1)\Big). \tag{2.5.4}$$

For a given empirical neighbourhood measure $\nu_n \in \mathcal{K}^{(n)}(\omega_n, \varpi_n)$ the numerator is probably too tricky to find explicitly.

However, an easy upper bound is

$$\binom{n}{n\nu_n(a,\ell),\, a \in \mathcal{X}, \ell \in \mathcal{N}(\mathcal{X})} \prod_{(a,b)} \binom{n\varpi_n(a,b)}{\ell_a^{(j)}(b),\, j = 1, \ldots, n\omega_n(a)} 2^{-\frac{n}{2}\varpi_n(\Delta)}, \tag{2.5.5}$$

where $\ell_a^{(j)}(b)$, $j = 1, \ldots, n\omega_n(a)$ are any enumeration of the family containing each $\ell(b)$ with multiplicity $n\nu_n(a,\ell)$. This upper bound is obtained by attaching edge-numbers without discounting for the possibility of multiple edges or loops. In the case $a = b$ initially edges are considered to be oriented and then the orientation is forgotten, leading to the extra term

$$2^{-\frac{n}{2}\varpi_n(\Delta)}.$$

Combining (2.5.3), (2.5.4), and (2.5.5) we get

$$\begin{aligned}
&\mathbb{P}\{M = \nu_n \,|\, \Phi(M) = (\omega_n, \varpi_n)\} \\
&\quad \le \prod_{a \in \mathcal{X}} \binom{n\omega_n(a)}{n\nu_n(a,\ell),\, \ell \in \mathcal{N}(\mathcal{X})} \prod_{(a,b)} \binom{n\varpi_n(a,b)}{\ell_a^{(j)}(b),\, j = 1, \ldots, n\omega_n(a)} \\
&\quad \times 2^{-\frac{n}{2}\varpi_n(\Delta)} \prod_{\{a,b\}} \prod_{k=1}^{n(a,b)} \Big(\tfrac{n^2\omega_n(a)\omega_n(b) - n\omega_n(a)\mathbb{1}_{\{a=b\}}}{1 + \mathbb{1}_{\{a=b\}}} - (k-1)\Big)^{-1}.
\end{aligned} \tag{2.5.6}$$

It remains to analyze the asymptotics of the upper bound. Using Stirling's formula, we obtain

$$\begin{aligned}
&\prod_{a\in\mathcal{X}} \binom{n\omega_n(a)}{n\nu_n(a,\ell),\, \ell\in\mathcal{N}(\mathcal{X})} \\
&\le \exp\Big(n\sum_a \omega_n(a)\log\omega_n(a) - n\sum_{(a,\ell)}\nu_n(a,\ell)\log\nu_n(a,\ell) + \tfrac{1}{n}\sum_a \tfrac{1}{12\omega_n(a)}\Big) \\
&\qquad \times \exp\big(\tfrac{m}{2}\log(2\pi n)\big).
\end{aligned}$$

We observe that

$$\prod_{j=1}^{n\omega_n(a)} (\ell_a^{(j)}(b))! = \exp\Big(n\sum_\ell \log\big(\ell(b)!\big)\,\nu_n(a,\,\ell)\Big),$$

and hence we have that

$$\begin{aligned}
&\binom{n\varpi_n(a,b)}{\ell_a^{(j)}(b),\, j\le n\omega_n(a)} \\
&\le \exp\Big(-n\sum_\ell \log\big(\ell(b)!\big)\,\nu_n(a,\ell) + n\varpi_n(a,b)\log\big(n\varpi_n(a,b)\big) - n\varpi_n(a,b)\Big) \\
&\qquad \times \exp\Big(\tfrac{1}{12n\varpi_n(a,b)} + \tfrac{1}{2}\log(2\pi n)\Big).
\end{aligned}$$

Next, we obtain,

$$\begin{aligned}
&\prod_{k=1}^{n(a,b)} \big(\tfrac{n^2\omega_n(a)\omega_n(b) - n\omega_n(a)\mathbb{1}_{\{a=b\}}}{1+\mathbb{1}_{\{a=b\}}}\big) - (k-1)\big) \\
&\ge \exp\Big(n(a,b)\,\log\big(\tfrac{n^2\omega_n(a)\omega_n(b)}{1+\mathbb{1}_{\{a=b\}}}\Big) \times \exp\Big(n(a,b)\,\log\big(1 - \tfrac{\mathbb{1}\{a=b\}}{2n\omega_n(a)} - \tfrac{n(a,b)}{n^2\omega_n(a)\omega_n(b)}\big)\Big).
\end{aligned}$$

Putting everything together, recalling that $H(\omega)$ is the entropy of a discrete measure ω, we get

$$\begin{aligned}
&\mathbb{P}\big\{M=\nu_n \,\big|\, \Phi(M) = (\omega_n,\varpi_n)\big\} \\
&\le \exp\Big(-nH(\omega_n) + nH(\nu_n) - n\sum_{(a,b)}\sum_\ell(\log\ell(b)!)\,\nu_n(a,\ell) - n\sum_{(a,b)}\varpi_n(a,b) \\
&+ n\sum_{(a,b)}\varpi_n(a,b)\log\varpi_n(a,b) - \tfrac{n}{2}\sum_{(a,b)}\varpi_n(a,b)\,\log\big((1+\mathbb{1}_{\{a=b\}})\,\omega_n(a)\omega_n(b)\big)\Big) \\
&\qquad \times \exp\Big(-\varpi_n(\Delta)\log 2 + \varepsilon_1^{(n)}\Big),
\end{aligned}$$

for a sequence $\varepsilon_1^{(n)}$ which does not depend on ν_n and satisfies $\lim_{n\uparrow\infty}\frac{1}{n}\varepsilon_1^{(n)}=0$. To give the right hand side the form as stated in the theorem, we observe that

$$\begin{aligned}
&H(\omega_n)-H(\nu_n)+\sum_{(a,b)}\sum_{\ell}(\log\ell(b)!)\,\nu_n(a,\ell)-\sum_{(a,b)}\varpi_n(a,b)\log\varpi_n(a,b)\\
&\qquad+\sum_{(a,b)}\varpi_n(a,b)+\sum_{(a,b)}\varpi_n(a,b)\,\log\omega_n(a)\\
&=\sum_{(a,\ell)}\nu_n(a,\ell)\Big[\log\tfrac{\nu_n(a,\ell)}{\omega_n(a)}-\sum_b\Big(\log\big(\tfrac{\varpi_n(a,b)}{\omega_n(a)}\big)^{\ell(b)}-\tfrac{\varpi_n(a,b)}{\omega_n(a)}-(\log\ell(b)!)\Big)\Big]\\
&=\sum_{(a,\ell)}\nu_n(a,\ell)\Big[\log\nu_n(a,\ell)-\log\Big(\omega_n(a)\prod_b\tfrac{(\varpi_n(a,b)/\omega_n(a))^{\ell(b)}\exp(-\varpi_n(a,b)/\omega_n(a))}{\ell(b)!}\Big)\Big]\\
&=H(\nu_n\,\|\,Q_n),
\end{aligned}$$

which completes the proof of Lemma 2.5.6. ∎

We can now complete the proof of the upper bound in Theorem 2.5.1 by combining Lemma 2.5.4 and Lemma 2.5.6. Suppose that $\Gamma\subset\mathcal{M}(\mathcal{X}\times\mathcal{N}(\mathcal{X}))$ is a closed set. Then,

$$\begin{aligned}
\mathbb{P}\Big\{M\in\Gamma\,\big|\,\Phi(M)=(\omega_n,\varpi_n)\Big\}&=\sum_{\nu_n\in\Gamma\cap\mathcal{K}^{(n)}(\omega_n,\varpi_n)}\mathbb{P}\{M=\nu_n\,|\,\Phi(M)=(\omega_n,\varpi_n)\}\\
&\le\sharp\mathcal{K}^{(n)}(\omega_n,\varpi_n)\,\exp\Big(-n\inf_{\nu_n\in\Gamma\cap\mathcal{K}^{(n)}(\omega_n,\varpi_n)}H(\nu_n\,\|\,Q_n)+\sup_{\nu_n\in\mathcal{K}^{(n)}(\omega_n,\varpi_n)}\varepsilon_1^{(n)}(\nu_n)\Big).
\end{aligned}$$

We have already seen that $\frac{1}{n}\sup_{\nu_n}\varepsilon_1^{(n)}(\nu_n)$ and $\frac{1}{n}\log\sharp\mathcal{K}^{(n)}(\omega_n,\varpi_n)$ converge to zero. It remains to check that

$$\lim_{n\to\infty}\sup_{\nu_n\in\mathcal{K}^{(n)}(\omega_n,\varpi_n)}\big|H(\nu_n\,\|\,Q_n)-H(\nu_n\,\|\,Q)\big|=0. \tag{2.5.7}$$

To do this, we observe that

$$\begin{aligned}
&H(\nu_n\,\|\,Q_n)-H(\nu_n\,\|\,Q)=\sum_{(a,\ell)\in\mathcal{X}\times\mathcal{N}(\mathcal{X})}\nu_n(a,\ell)\log\tfrac{Q(a,\ell)}{Q_n(a,\ell)}\\
&=-H(\omega_n\,\|\,\omega)-H(\varpi_n\,\|\,\varpi)-\sum_{a,b\in\mathcal{X}}\varpi(a,b)\tfrac{\omega_n(a)}{\omega(a)}+\sum_{a,b\in\mathcal{X}}\varpi(a,b)\log\tfrac{\omega_n(a)}{\omega(a)}+\|\varpi_n\|.
\end{aligned} \tag{2.5.8}$$

Note that this expression does not depend on ν_n. As the first, second and fourth term of (2.5.8) converge to 0, and the third and fifth term converge to $\|\varpi\|$, the expression (2.5.8) vanishes in the limit, and this completes the proof of the upper bound in Theorem 2.5.1.

2.5.5 An Upper Bound on the Support of Empirical Neighbourhood Measures

The support, denoted $\mathcal{S}(\nu) \subset \mathcal{X} \times \mathcal{N}(\mathcal{X})$, of an empirical neighbourhood measure ν of a graph with n vertices is naturally bounded by n. For the proof of the lower bound in Theorem 2.5.1 we need a better upper bound. Again we abbreviate $m := |\mathcal{X}|$ and let

$$C := 2^m \frac{\Gamma(m+2)^{\frac{m}{m+1}}}{\Gamma(m)} \text{ and } D := 2^m \frac{(m+1)^m}{\Gamma(m)},$$

where $\Gamma(\cdot)$ is the Gamma function.

Lemma 2.5.7. *For every $(\omega_n, \varpi_n) \in \mathcal{M}_n(\mathcal{X}) \times \tilde{\mathcal{M}}_n(\mathcal{X} \times \mathcal{X})$ and $\nu_n \in \mathcal{M}_n(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$ with $\Phi(\nu_n) = (\omega_n, \varpi_n)$, we have*

$$\sharp\mathcal{S}(\nu_n) \leq C \left[n\|\varpi_n\|\right]^{\frac{m}{m+1}} + D. \tag{2.5.9}$$

The following lemma provides a step in the proof of Lemma 2.5.7.

Lemma 2.5.8. *Suppose $j \in \mathbb{N} \cup \{0\}$ and $n \in \mathbb{N}$. Then,*

$$\tfrac{1}{\Gamma(n)}\, j^{n-1} \leq \sharp\Big\{(l_1, \ldots, l_n) \in \big(\mathbb{N} \cup \{0\}\big)^n \colon l_1 + \ldots + l_n = j\Big\} \leq \tfrac{1}{\Gamma(n)}\,(j+n)^{n-1}. \tag{2.5.10}$$

Proof. The proof is by induction on n. Equation (2.5.10) holds trivially for all $j \in \mathbb{N} \cup \{0\}$ and $n = 1, 2$, so we assume it holds for all j and $n \geq 2$.
By the induction hypothesis, for any j,

$$\begin{aligned}
\frac{1}{\Gamma(n)} \sum_{l=0}^{j} (j-l)^{n-1} &\leq \sum_{l=0}^{j} \sharp\Big\{(l_1, \ldots, l_{n-1}) \in \big(\mathbb{N} \cup \{0\}\big)^{n-1} \colon l_1 + \ldots + l_{n-1} = j - l\Big\} \\
&= \sharp\Big\{(l_1, \ldots, l_n) \in \big(\mathbb{N} \cup \{0\}\big)^n \colon l_1 + \ldots + l_n = j\Big\} \\
&\leq \frac{1}{\Gamma(n)} \sum_{l=0}^{j} (j - l + n)^{n-1}.
\end{aligned}$$

For the first and last term, we obtain the lower and upper bounds

$$\sum_{l=0}^{j}(j-l)^{n-1} \geq \int_0^j y^{n-1}\,dy = \frac{1}{n}\,j^n = \frac{\Gamma(n)}{\Gamma(n+1)}j^n$$

and

$$\begin{aligned}\sum_{l=0}^{j}(j-l+n)^{n-1} &\leq \int_n^{j+n} y^{n-1}dy \\ &\leq \int_0^{j+n+1} y^{n-1}dy = \frac{1}{n}\,(j+n+1)^n = \frac{\Gamma(n)}{\Gamma(n+1)}(j+n+1)^n,\end{aligned}$$

which yields inequality (2.5.10) for $n+1$ instead of n, and completes the induction. ∎

Proof of Lemma 2.5.7. Suppose $(\omega_n, \varpi_n) \in \mathcal{M}_n(\mathcal{X}) \times \mathcal{M}_n(\mathcal{X} \times \mathcal{X})$. Let

$$\begin{aligned}a_m(j) &:= \sharp\{(a,\ell) \in \mathcal{X} \times \mathcal{N}(\mathcal{X}) \,:\, \sum_{b\in\mathcal{X}} \ell(b) = j\} \\ &= m \times \sharp\Big\{(l_1,\,\ldots,\,l_m) \in \big(\mathbb{N}\cup\{0\}\big)^m \colon l_1+\ldots+l_m = j\Big\}.\end{aligned}$$

For any positive integer k we write

$$\theta_k = \min\Big\{\theta \in \mathbb{N} : \sum_{j=0}^{\theta} a_m(j) \geq k\Big\}.$$

We observe from Lemma 2.5.8 that,

$$k \leq \sum_{j=0}^{\theta_k} a_m(j) \leq m \sum_{j=0}^{\theta_k} \tfrac{1}{\Gamma(m)}(j+m)^{m-1} \leq \tfrac{m}{\Gamma(m)} \int_0^{\theta_k+m} y^{m-1}dy = \tfrac{1}{\Gamma(m)}\big(\theta_k+m\big)^m.$$

Thus, we have $\theta_k \geq \big(k\Gamma(m)\big)^{\frac{1}{m}} - m =: \alpha_k$. This yields

$$\sum_{j=0}^{\theta_k} j a_m(j) \geq \tfrac{1}{\Gamma(m)} \sum_{j=0}^{\lceil\alpha_k\rceil} j^m \geq \tfrac{1}{\Gamma(m)} \int_0^{\alpha_k-1} y^m dy \geq \tfrac{1}{\Gamma(m+2)}\big(\alpha_k - 1\big)^{m+1}, \qquad (2.5.11)$$

where $\lceil y \rceil$ is the smallest integer greater or equal to y.

Observe that the size of the support of the measure $\nu_n \in \mathcal{K}^{(n)}(\omega_n, \varpi_n)$ satisfies

$$\sharp\mathcal{S}(\nu_n) \le \max\Big\{k\colon \sum_{j=0}^{\theta_k} j a_m(j) \le n\|\varpi_n\|\Big\},$$

and hence, using (2.5.11) and the inequality $(a+b)^m \le 2^m(a^m+b^m)$ for $a,b\ge 0$,

$$\begin{aligned}\sharp\mathcal{S}(\nu_n) &\le \max\Big\{k\colon \tfrac{1}{\Gamma(m+2)}\big(\alpha_k-1\big)^{m+1} \le n\|\varpi_n\|\Big\}\\ &\le \Gamma(m)^{-1}\Big((n\|\varpi_n\|)^{\frac{1}{m+1}}\Gamma(m+2)^{\frac{1}{m+1}}+m+1\Big)^m \le C\,(n\|\varpi_n\|)^{\frac{m}{m+1}}+D\,,\end{aligned}$$

where the constants C, D were define before the formulation of the lemma. ■

2.5.6 Approximation by Empirical Neighbourhood Measures

Throughout this section we assume that $\omega_n \in \mathcal{M}_n(\mathcal{X})$ with $\omega_n \to \omega$, $\varpi_n \in \tilde{\mathcal{M}}_n(\mathcal{X}\times\mathcal{X})$ with $\varpi_n \to \varpi$, and that $\nu \in \mathfrak{M}[\omega,\varpi]$. Our aim is to show that ν can be approximated in the weak topology by the empirical neighbourhood measure of a graph with n vertices, empirical colour measure ω_n, empirical pair measure ν_n, and the additional feature that the degree of any vertex is bounded by $n^{1/3}$. The approximation will be done in four steps.

We denote by d the metric of total variation, i.e.

$$d(\nu,\tilde\nu) = \tfrac12 \sum_{(a,\ell)\in\mathcal{X}\times\mathcal{N}(\mathcal{X})} |\nu(a,\ell)-\tilde\nu(a,\ell)|, \qquad \text{for } \nu,\tilde\nu \in \mathcal{M}(\mathcal{X}\times\mathcal{N}(\mathcal{X})).$$

This metric generates the weak topology.

Lemma 2.5.9 (Approximation Step 1). *For every $\varepsilon>0$, there exist $\hat\nu \in \mathcal{M}(\mathcal{X}\times\mathcal{N}(\mathcal{X}))$ and $\hat\varpi \in \tilde{\mathcal{M}}_*(\mathcal{X}\times\mathcal{X})$ such that $|\varpi(a,b)-\hat\varpi(a,b)|\le\varepsilon$ for all $a,b\in\mathcal{X}$, $d(\nu,\hat\nu)\le\varepsilon$ and $(\hat\varpi,\hat\nu)$ is consistent.*

Proof

By our assumption (ϖ,ν) is sub-consistent. For any $b\in\mathcal{X}$ define $e^{(b)}\in\mathcal{N}(\mathcal{X})$ by $e^{(b)}(a)=0$ if $a\neq b$, and $e^{(b)}(b)=1$. For large n define measures $\hat\nu_n \in \mathcal{M}(\mathcal{X}\times\mathcal{N}(\mathcal{X}))$ by

$$\hat\nu_n(a,\ell) = \nu(a,\ell)\Big(1-\tfrac{\|\varpi\|-\|\langle\nu(\cdot,\ell),\ell(\cdot)\rangle\|}{n}\Big) + \sum_{b\in\mathcal{X}} \mathbb{1}\{\ell = ne^{(b)}\}\,\tfrac{\varpi(a,b)-\langle\nu(\cdot,\ell),\ell(\cdot)\rangle(a,b)}{n}\,.$$

Note that $\hat{\nu}_n \to \nu$ and that, for all $a, b \in \mathcal{X}$,

$$\begin{aligned}\sum_{\ell \in \mathcal{N}(\mathcal{X})} \hat{\nu}_n(a,\ell)\ell(b) &= \Big(1 - \tfrac{\|\varpi\| - \|\langle \nu(\cdot,\ell), \ell(\cdot)\rangle\|}{n}\Big) \sum_{\ell \in \mathcal{N}(\mathcal{X})} \nu(a,\ell)\ell(b) + \varpi(a,b) \\ &\qquad\qquad\qquad\qquad\qquad - \langle \nu(\cdot,\ell), \ell(\cdot)\rangle(a,b) \\ &= \varpi(a,b) - \tfrac{\|\varpi\| - \|\langle \nu(\cdot,\ell), \ell(\cdot)\rangle\|}{n} \langle \nu(\cdot,\ell), \ell(\cdot)\rangle(a,b) \stackrel{n\uparrow\infty}{\longrightarrow} \varpi(a,b)\,.\end{aligned}$$

Hence, defining $\hat{\varpi}_n$ by $\hat{\varpi}_n(a,b) = \sum \hat{\nu}_n(a,\ell)\ell(b)$, we have a sequence of consistent pairs $(\hat{\varpi}_n, \hat{\nu}_n)$ converging to (ϖ, ν), as required.

∎

Lemma 2.5.10 (Approximation Step 2). *For every $\varepsilon > 0$, there exists $n(\varepsilon)$ such that, for all $n \geq n(\varepsilon)$ there exists $\nu_n \in \mathcal{M}_n(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$ with $\Phi(\nu_n) = (\omega_n, \varpi_n)$ such that $d(\nu_n, \nu) \leq \varepsilon$.*

The key to the construction of the measure ν_n is the following 'law of large numbers'.

Lemma 2.5.11. *For every $\delta > 0$, there exists $\hat{\nu} \in \mathcal{M}(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$ with $d(\nu, \hat{\nu}) < \delta$ such that, for i.i.d. $\mathcal{N}(\mathcal{X})$-valued random variables ℓ_j^a, $j = 1, \ldots, n\omega_n(a)$ with law $\hat{\nu}(\,\cdot\,|\,a) := \hat{\nu}(\{a\} \times \,\cdot\,)/\hat{\nu}_1(a)$, almost surely,*

$$\limsup_{n\to\infty} \Big(\frac{1}{n} \sum_{j=1}^{n\omega_n(a)} \ell_j^a(b) - \varpi_n(a,b)\Big) \leq \delta, \qquad \forall a, b \in \mathcal{X}. \tag{2.5.12}$$

Proof. As $\nu \in \mathfrak{M}[\omega, \varpi]$ we can use Lemma 2.5.9 to choose $(\hat{\varpi}, \hat{\nu})$ consistent such that $d(\nu, \hat{\nu}) < \delta$ and, for all $a, b \in \mathcal{X}$,

$$\tfrac{\nu_1(a)}{\hat{\nu}_1(a)} \leq 1 + \tfrac{\delta}{\|\varpi\|+1} \quad \text{and} \quad \hat{\varpi}(a,b)\big(1 + \tfrac{\delta}{\|\varpi\|+1}\big) \leq \varpi(a,b)\big(1 + \tfrac{\delta}{\varpi(a,b)}\big)\,.$$

The random variables $\ell_j^a(b)$, $j = 1, \ldots, n\omega_n(a)$ are i.i.d. with expectation

$$\mathbb{E}\ell_1^a(b) = \sum_{\ell} \ell(b)\hat{\nu}(\ell\,|\,a) = \frac{\hat{\varpi}(a,b)}{\hat{\nu}_1(a)}\,.$$

Hence, by the strong law of large numbers, almost surely,

$$\limsup_{n\to\infty} \Big(\frac{1}{n} \sum \ell_j^a(b) - \varpi_n(a,b)\Big) \leq \tfrac{\nu_1(a)}{\hat{\nu}_1(a)}\hat{\varpi}(a,b) - \varpi(a,b) \leq \delta,$$

$$\limsup_{n\to\infty} \Big(\frac{1}{n} \sum_{j=1} \ell_j^a(b) - \varpi_n(a,b)\Big) \leq \tfrac{\nu_1(a)}{\hat{\nu}_1(a)}\hat{\varpi}(a,b) - \varpi(a,b) \leq \delta,$$

where we also used that $\omega_n(a) \to \omega(a) = \nu_1(a)$ and $\varpi_n(a,b) \to \varpi(a,b)$. ■

Proof of Lemma 2.5.10. We use a random construction. Given $\nu \in \mathfrak{M}[\omega, \varpi]$ and $\varepsilon > 0$, choose $\hat{\nu}$ as in Lemma 2.5.11 with $\delta = \varepsilon/(3m)$, where m is the cardinality of $\mathcal{X}$. For every $a \in \mathcal{X}$, we draw tuples ℓ_j^a, $j = 1, \ldots, n\omega_n(a)$ independently according to $\hat{\nu}(\,\cdot\,|\,a)$ and define $e_n(a,b)$ by

$$e_n(a,b) := \frac{1}{n} \sum_{j=1}^{n\omega_n(a)} \ell_j^a(b) - \varpi_n(a,b), \quad \text{for all } a, b \in \mathcal{X}.$$

We modify the tuples $(\ell_j^a : j = 1, \ldots, n\omega_n(a))$ as follows:

- If $e_n(a,b) < 0$, we add an amount to the last element $\ell^a_{n\omega_n(a)}(b)$ such that the modified tuple satisfies $e_n(a,b) = 0$;
- if $e_n(a,b) > 0$, by Lemma 2.5.11, the 'overshoot' $ne_n(a,b)$ cannot exceed $n\delta$. We successively deduct one from the nonzero elements in $\ell_j^a(b)$, $j = 1, \ldots, n\omega_n(a)$ until the modified tuples satisfy $e_n(a,b) = 0$;
- if $e_n(a,b) = 0$ we do not modify $\ell_j^a(b)$.

We denote by $(\tilde{\ell}_j^a : j = 1, \ldots, n\omega_n(a))$ the tuples after all modifications.

For each $a \in \mathcal{X}$ define probability measures $\tilde{\Delta}_n(\,\cdot\,|\,a)$ and $\Delta_n(\,\cdot\,|\,a)$ by

$$\tilde{\Delta}_n(\ell\,|\,a) = \frac{1}{n\omega_n(a)} \sum_{j=1}^{n\omega_n(a)} \mathbb{1}_{\{\tilde{\ell}_j^a = \ell\}}, \qquad \text{for } \ell \in \mathcal{N}(\mathcal{X}),$$

respectively,

$$\Delta_n(\ell\,|\,a) = \frac{1}{n\omega_n(a)} \sum_{j=1}^{n\omega_n(a)} \mathbb{1}_{\{\ell_j^a = \ell\}}, \qquad \text{for } \ell \in \mathcal{N}(\mathcal{X}).$$

We define probability measures $\tilde{\nu}_n \in \mathcal{M}_n(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$ and $\nu_n \in \mathcal{M}_n(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$ by $\tilde{\nu}_n(a,\ell) = \omega_n(a)\Delta_n(\ell\,|\,a)$, respectively $\nu_n(a,\ell) = \omega_n(a)\tilde{\Delta}_n(\ell\,|\,a)$, for $(a,\ell) \in \mathcal{X} \times \mathcal{N}(\mathcal{X})$. Recall from our modification procedure that, in the worst case, we have changed $nm\delta$ of the tuples. Thus,

$$d(\tilde{\nu}_n, \nu_n) \le m\delta \le \tfrac{1}{3}\varepsilon\,.$$

As a result of our modifications we have $\Phi(\nu_n) = (\omega_n, \varpi_n)$. We observe that, for all $(a,\ell) \in \mathcal{X} \times \mathcal{N}(\mathcal{X})$, the random variables

$$\mathbb{1}\{\ell_1^a = \ell\}, \ldots, \mathbb{1}\{\ell^a_{n\omega_n(a)} = \ell\}$$

are independent Bernoulli random variables with success probability $\hat{\nu}(\ell \,|\, a)$ and hence, almost surely,

$$\lim_{n\to\infty} \Delta_n(\ell \,|\, a) = \hat{\nu}(\ell \,|\, a)\,.$$

Therefore, for all $(a,\ell) \in \mathcal{X} \times \mathcal{N}(\mathcal{X})$, we obtain $\lim_{n\to\infty} \tilde{\nu}_n(a,\ell) = \hat{\nu}(a,\ell)$, almost surely. Thus, almost surely, for all large n, we have $d(\nu_n,\nu) \le d(\nu_n,\tilde{\nu}_n) + d(\tilde{\nu}_n,\hat{\nu}) + d(\hat{\nu},\nu) \le \varepsilon$, as claimed. ∎

Lemma 2.5.12 (Approximation Step 3). *Let $\nu_n \in \mathcal{M}_n(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$ with $\Phi(\nu_n) = (\omega_n, \varpi_n)$. For every $\varepsilon > 0$ there exists $n(\varepsilon)$ such that, for all $n \ge n(\varepsilon)$, we can find $\tilde{\nu}_n \in \mathcal{M}_n(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$ with $\Phi(\tilde{\nu}_n) = (\omega_n, \varpi_n)$ and $d(\nu_n, \tilde{\nu}_n) < \varepsilon$, such that*

$$\sum_{b\in\mathcal{X}} \ell(b) \le n^{1/3} \text{ for } \tilde{\nu}_n\text{-almost every } (a,\ell). \tag{2.5.13}$$

Proof. As $\nu_n \in \mathcal{M}_n(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$, there is a representation

$$\nu_n = \frac{1}{n}\sum_{k=1}^{n} \delta_{(a_k,\ell_k)}, \qquad \text{for } a_k \in \mathcal{X}, \ell_k \in \mathcal{N}(\mathcal{X}).$$

Fix $\delta > 0$ and $a \in \mathcal{X}$. Look at the sets

- $V^+ = \{1 \le k \le n \colon a_k = a,\ \sum_b \ell_k(b) > n^{1/3}\}$ with $\sharp V^+ \le (n \sum_b \varpi_n(a,b))^{2/3}$,
- $V^- = \{1 \le k \le n \colon a_k = a,\ \sum_b \ell_k(b) \le n^{1/4}\}$ with $\sharp V^- \ge n - (n \sum_b \varpi_n(a,b))^{3/4}$.

For each $k \in V^+$ we replace ℓ_k by a smaller vector $\tilde{\ell}_k$ such that $\sum_b \tilde{\ell}_k(b) = n^{1/3}$. As

$$\sum_{k\in V_1}\sum_b \ell_k(b) \le n \sum_{b\in\mathcal{X}} \varpi_n(a,b)$$

we replace (for large n) no more than δn of the vectors ℓ_k, $k \in V^-$, by larger vectors $\tilde{\ell}_k$ such that

$$\sum_b \tilde{\ell}_k(b) \le n^{1/3} \text{ and } \sum_{k=1}^{n}\sum_{b\in\mathcal{X}} \mathbb{1}\{a_k = a\}\, \tilde{\ell}_k(b) = \sum_{k=1}^{n}\sum_{b\in\mathcal{X}} \mathbb{1}\{a_k = a\}\, \ell_k(b)\,,$$

where we use the convention $\tilde{\ell}_k = \ell_k$ if this vector was not changed in the procedure. Performing such an operation for every $a \in \mathcal{X}$ we may define

$$\tilde{\nu}_n = \frac{1}{n} \sum_{k=1}^{n} \delta_{(a_k, \tilde{\ell}_k)},$$

and observe that (2.5.13) holds and $\Phi(\tilde{\nu}_n) = (\omega_n, \varpi_n)$. Moreover,

$$d(\nu_n, \tilde{\nu}_n) \le \sharp\mathcal{X} \, \tfrac{1}{2n}\big((n \sum_b \varpi_n(a,b))^{2/3} + \delta n\big),$$

which is less than $\varepsilon > 0$ for a suitable choice of $\delta > 0$, and all sufficiently large n.
■

Lemma 2.5.13 (Approximation Step 4). *Let $\nu_n \in \mathcal{M}_n(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$ with $\Phi(\nu_n) = (\omega_n, \varpi_n)$. For every $\varepsilon > 0$ there exists $n(\varepsilon)$ such that, for all $n \ge n(\varepsilon)$, we can find $\tilde{\nu}_n \in \mathcal{K}^{(n)}(\omega_n, \varpi_n)$ with $d(\nu_n, \tilde{\nu}_n) < \varepsilon$ such that* (2.5.13) *holds.*

Proof. By Lemma 2.5.12 we may assume that ν_n can be represented as

$$\nu_n = \frac{1}{n} \sum_{k=1}^{n} \delta_{(a_k, \ell_k)}, \qquad \text{for } a_k \in \mathcal{X}, \ell_k \in \mathcal{N}(\mathcal{X}) \text{ with } \sum_{b \in \mathcal{X}} \ell_k(b) \le n^{1/3}.$$

For the proof it suffices to construct, for every $\varepsilon > 0$ and all large n, a (coloured) random graph X with n vertices such that

$$\limsup_{n \to \infty} \mathbb{P}\big\{ d(\nu_n, M(X)) \ge \varepsilon \big\} = 0, \tag{2.5.14}$$

We now describe the random procedure that generates X. First, equip each vertex with an element of $\mathcal{X} \times \mathcal{N}(\mathcal{X})$ by drawing without replacement from the collection $\{(a_1, \ell_1), \dots, (a_n, \ell_n)\}$. We denote by $V(a)$ the collection of vertices which have colour $a \in \mathcal{X}$, and observe that $\sharp V(a) = n\omega_n(a)$.

Now fix $a, b \in \mathcal{X}$. We construct two collections of vertices: If a vertex is equipped with (a_k, ℓ_k) and $a_k = a$, then it is represented $\ell_k(b)$ times in the first collection, $W(a)$. If $a_k = b$, then it is represented $\ell_k(b)$ times in the second collection $W(b)$. Hence there are exactly $n\varpi(a,b)/2$ vertices of colour a in collection $W(a)$, and exactly $n\varpi(a,b)/2$ vertices of colour b in collection $W(b)$.

We now match vertices from the two collections randomly: At each step $k = 1, \dots, n\varpi_n(a,b)/2$, we randomly pick two vertices $V_1^k \in W(a)$ and

$V_2^k \in W(b)$. We connect V_1^k and V_2^k by an edge unless $V_1^k = V_2^k$ or the two vertices are already connected. If one of these two things happen, then we simply choose an edge randomly from the set of all possible edges connecting colours a and b, which are not yet present in the graph and whose introduction does not violate (2.5.13).

This completes the construction of a graph with $L^1(X) = \omega_n$, $L^2(X) = \varpi_n$ and

$$d(\nu_n\,,\,M(X)) \leq \tfrac{2}{n} \sum_{a,b\in\mathcal{X}} B^n(a,b)\,, \tag{2.5.15}$$

where $B^n(a,b)$ is the total number of steps $k \in \{1,\dots,n\varpi_n(a,b)/2\}$ at which there is disparity between the vertices V_1^k, V_2^k drawn and the vertices which formed the k^{th} edge connecting a and b in the random graph construction.

Given $a, b \in \mathcal{X}$, Using condition (2.5.13), the probability that $V_1^k = V_2^k$ or the two vertices are already connected is less or equal

$$p_{[k]}(a,b) := \tfrac{2n^{1/3}}{n\varpi_n(a,b)} 1\!\!1_{\{a=b\}} + \big(1 - \tfrac{2n^{1/3}}{n\varpi_n(a,b)} 1\!\!1_{\{a=b\}}\big) \tfrac{4(k-1)n^{2/3}}{(n\varpi_n(a,b))^2}.$$

$B^n(a,b)$ is a sum of independent Bernoulli random variables X_1, ..., $X_{n\varpi_n(a,b)/2}$ with 'success' probabilities less or equal $p_{[1]}(a,b),\dots,p_{[n\varpi_n(a,b)/2]}(a,b)$. So its expectation satisfies

$$\mathbb{E}B^n(a,b) \leq \sum_{k=1}^{n\varpi_n(a,b)/2} kp_{[k]}(a,b) = n^{1/3} 1\!\!1_{\{a=b\}} + n^{2/3}\big(1 - \tfrac{2}{n^{2/3}\varpi_n(a,b)}\big)\big(1 - \tfrac{2}{n\varpi_n(a,b)}\big).$$

Hence, we have that $\mathrm{Var}(B^n(a,b)) \leq \mathbb{E}B^n(a,b) = o(n)$. By Bernstein's inequality, see for example [BBL04, Theorem 3], we obtain, for any $\epsilon > 0$,

$$\mathbb{P}\big\{B^n(a,b) \geq \mathbb{E}B^n(a,b) + 2n\epsilon\big\} \leq \exp\Big\{ - \tfrac{n^2\epsilon^2}{2\,(\mathrm{Var}(B^n(a,b))+n\epsilon/3)}\Big\}\,.$$

Using the bounds on $\mathbb{E}B^n(a,b)$ and $\mathrm{Var}(B^n(a,b))$, we obtain, for sufficiently large n,

$$\mathbb{P}\big\{B^n(a,b) \geq o(n) + 2n\epsilon\big\} \leq \exp\Big\{ - \tfrac{n^2\epsilon^2}{2o(n)+2n\epsilon/3}\Big\}\,.$$

Now suppose a small $\varepsilon > 0$ and large $A > 0$ are given. Let $\delta = \varepsilon/(2m^2)$. Suppose that $B^n(a,b) \leq n\delta$, for all $a,b \in \mathcal{X}$. Then, by (2.5.15), $d(M(X),\nu_n) \leq 2\delta m^2 = \varepsilon$.

Hence,

$$\begin{aligned}\mathbb{P}\{d(M(X),\nu_n) > \varepsilon\} &\le \sum_{a,b\in\mathcal{X}} \mathbb{P}\{B^n(a,b) \ge n\delta\} \\ &\le m^2 \sup_{a,b\in\mathcal{X}} \mathbb{P}\{B^n(a,b) \ge o(n) + (n\delta)/2\} \\ &\le m^2 \exp\Big\{ - \tfrac{n^2\delta^2}{8o(n)+(4n\delta)/3}\Big\}.\end{aligned}$$

This completes the proof of the lemma. ■

2.5.7 Lower Bound in the Second Auxiliary Principle

There is a partial analogue to Lemma 2.5.6 for the lower bounds.

Lemma 2.5.14. *For any sequence (ν_n) with $\nu_n \in \mathcal{K}^{(n)}(\omega_n, \varpi_n)$, which satisfies* (2.5.13), *and $\varepsilon > 0$, we have*

$$\mathbb{P}\{d(M,\nu_n) < \varepsilon \,|\, \Phi(M) = (\omega_n, \varpi_n)\} \ge \exp\big(- nH(\nu_n \,\|\, Q_n) - \varepsilon_2^{(n)}(\nu_n)\big)\,,$$

where Q_n is as Lemma 2.5.6 and

$$\lim_{n\uparrow\infty} \tfrac{1}{n}\, \varepsilon_2^{(n)}(\nu_n) = 0\,.$$

Proof of Lemma 2.5.14. We use the notation and some results from the proof of the upper bound, Lemma 2.5.6. Recall that

$$\mathbb{P}\{M = \nu_n \,|\, \Phi(M) = (\omega_n, \varpi_n)\} = \frac{\sharp\{\tilde{x} \in \mathcal{A}_n(\omega_n, \varpi_n) \,:\, M \circ \Psi(\tilde{x}) = \nu_n\}}{\sharp\{\tilde{x} \in \mathcal{A}_n(\omega_n, \varpi_n))\}}\,,$$

and that the denominator was evaluated in (2.5.4) as

$$\binom{n}{n\omega_n(a),\ a \in \mathcal{X}} \prod_{\{a,b\}} \prod_{k=1}^{n(a,b)} \big(\tfrac{n^2\omega_n(a)\omega_n(b) - n\omega_n(a)\mathbb{1}_{\{a=b\}}}{1+\mathbb{1}_{\{a=b\}}} - (k-1)\big)\,.$$

The numerator can be estimated from below by assigning edge-numbers to the vertices in a manner cautious to avoid loops and double edges. In each step the number of assignments, which lead to multiple edges or loops, is bounded by the square of the maximal vertex degree and hence by $n^{2/3}$. Hence the numerator is bounded from below by

$$\binom{n}{n\nu_n(a,\ell),\, a\in\mathcal{X},\ell\in\mathcal{N}(\mathcal{X})}\prod_{(a,b)}\frac{(n\varpi_n(a,b)-n^{2/3})!}{\prod_{j=1}^{n\omega_n(a)}(\ell_a^{(j)}(b))!}\,2^{-\frac{n}{2}\varpi_n(\Delta)}\,.$$

We again use Stirling's formula as in the proof of the upper bound. For the denominator we get the same main terms as in Lemma 2.5.6 with slightly different error terms, which however do not depend on ν_n.

More interestingly, we have

$$\begin{aligned}
&\prod_{a\in\mathcal{X}}\binom{n\omega_n(a)}{n\nu_n(a,\ell),\,\ell\in\mathcal{N}(\mathcal{X})}\\
&\geq\exp\Big(n\sum_a\omega_n(a)\log\omega_n(a)-n\sum_{(a,\ell)}\nu_n(a,\ell)\log\nu_n(a,\ell)-\tfrac{|S(\nu_n)|}{2}\log(2\pi n)\Big)\\
&\qquad\times\exp\Big(-\sum_{\substack{(a,\ell)\\ n\nu_n(a,\ell)\geq 1}}\tfrac{1}{12n\nu_n(a,\ell)}\Big)\,,
\end{aligned}$$

where the exponent in the error term is of order $o(n)$, by the bound on the size of the support of ν_n given in Lemma 2.5.7.

Further,

$$\begin{aligned}
&\frac{(n\varpi_n(a,b)-n^{2/3})!}{\prod_{j=1}^{n\omega_n(a)}(\ell_a^{(j)}(b))!}\\
&\geq\exp\Big(-n\sum_\ell\log\big(\ell(b)!\big)\,\nu_n(a,\ell)+n\varpi_n(a,b)\log\big(n\varpi_n(a,b)\big)-n\varpi_n(a,b)\Big)\\
&\qquad\qquad\times\exp\Big(-n^{2/3}\log\big(n\varpi_n(a,b)\big)+n\varpi_n(a,b)\,\log\big(1-\tfrac{n^{2/3}}{n\varpi_n(a,b)}\big)\Big)\,,
\end{aligned}$$

and the result follows by combining this with facts discussed in the context of the upper bound. ∎

To complete the proof of the lower bound in Theorem 2.5.1, take an open set $\Gamma\subset\mathcal{M}(\mathcal{X}\times\mathcal{N}(\mathcal{X}))$. Then, for any $\nu\in\Gamma\cap\mathfrak{M}[\omega,\varpi]$ we may find $\varepsilon>0$ with the ball around ν of radius $2\varepsilon>0$ contained in Γ. By our approximation, Lemma 2.5.13, we may find $\nu_n\in\Gamma\cap\mathcal{K}^{(n)}(\omega_n,\varpi_n)$ with $d(\nu_n,\nu)\downarrow 0$ and (ϖ_n,ν_n) sub-consistent, such that (2.5.13) holds. Hence, for all large $n\geq n(\varepsilon)$,

$$\begin{aligned}
\mathbb{P}\Big\{M\in\Gamma\,\big|\,\Phi(M)=(\omega_n,\varpi_n)\Big\}&\geq\mathbb{P}\big\{d(\nu_n,M)<\varepsilon\,\big|\,\Phi(M)=(\omega_n,\varpi_n)\big\}\\
&\geq\exp\Big(-nH(\nu_n\,\|\,Q_n)+\varepsilon_2^{(n)}(\nu_n)\Big).
\end{aligned}$$

We observe that

$$\lim_{n\to\infty} H(\nu_n \,\|\, Q_n) - H(\nu \,\|\, Q) = \lim_{n\to\infty} H(\nu_n \,\|\, Q_n) - H(\nu_n \,\|\, Q) + \lim_{n\to\infty} H(\nu_n \,\|\, Q) - H(\nu \,\|\, Q) = 0,$$

where the last term vanishes by continuity of the entropy, and the first term was shown to vanish in the proof of Lemma 2.5.14. As $\nu \in \Gamma \cap \mathfrak{M}[\omega, \varpi]$ was arbitrary, this completes the proof of Theorem 2.5.1.

2.6 The Main LDP by Method of Mixtures

We denote by $\Theta_n := \mathcal{M}_n(\mathcal{X}) \times \tilde{\mathcal{M}}_n(\mathcal{X} \times \mathcal{X})$ and $\Theta := \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X})$. With

$$P^{(n)}_{(\omega_n,\varpi_n)}(\nu_n) := \mathbb{P}\big\{M = \nu_n \,\big|\, \Phi(M) = (\omega_n, \varpi_n)\big\}\,,$$
$$P^{(n)}(\omega_n, \varpi_n) := \mathbb{P}\big\{(L^1, L^2) = (\omega_n, \varpi_n)\big\}$$

the joint distribution of L^1, L^2 and M is the mixture of $P^{(n)}_{(\omega_n,\varpi_n)}$ with $P^{(n)}(\omega_n, \varpi_n)$ defined as

$$d\tilde{P}^n(\omega_n, \varpi_n, \nu_n) := dP^{(n)}_{(\omega_n,\varpi_n)}(\nu_n)\, dP^{(n)}(\omega_n, \varpi_n)\,. \tag{2.6.1}$$

Biggins, see Theorem 1.3.4, gives criteria for the validity of large deviation principles for the mixtures and for the goodness of the rate function if individual large deviation principles are known. The following two lemmas ensure validity of these conditions.

Lemma 2.6.1. $(\tilde{P}^n\colon n \in \mathbb{N})$ *is exponentially tight.*

Proof. Given $k \in \mathbb{N}$, we observe from Lemma 2.4.9 that there exists $N(k) \in \mathbb{N}$ such that, for all sufficiently large n,

$$\mathbb{P}\big\{M(\{\|\ell\| \ge 2kN(k)\}) \ge k^{-1} \text{ or } \|L^2\| \ge 2N(k)\big\} \le \mathbb{P}\big\{|E| \ge nN(k)\big\} \le e^{-kn}.$$

Now, for any $\theta > 0$, we define the set Ξ_θ by

$$\Xi_\theta := \big\{(\varpi, \nu) \in \tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{N}(\mathcal{X})) \,:\nu\{\|\ell\| > 2lN(l)\} < l^{-1}\, \forall l \ge \theta \text{ and } \|\varpi\| < 2N(\theta)\big\}.$$

As $\{\|\ell\| \le 2lN(l)\} \subset \mathcal{N}(\mathcal{X})$ is finite, hence compact, the set Ξ_θ is relatively compact in the weak topology, by Prohorov's criterion. Moreover, we have that

$$\tilde{P}^{(n)}\big((\Xi_\theta)^c\big) \le \mathbb{P}\{\|L^2\| \ge 2N(\theta)\} + \sum_{l=\theta}^{\infty} \mathbb{P}\big\{M(\{\|\ell\| > 2lN(l)\}) \ge l^{-1}\big\} \le C(\theta)\, e^{-n\theta}.$$

Therefore, $\limsup_{n\to\infty} \frac{1}{n} \log \tilde{P}^n((\mathrm{cl}\, \Xi_\theta)^c)) \le -\theta$, which completes the proof, as $\theta > 0$ was arbitrary. ■

Recall that $\tilde{J}_{(\omega,\varpi)}$ is defined by

$$\tilde{J}_{(\omega,\varpi)}(\nu) = \begin{cases} H(\nu \,\|\, Q) & \text{if } \nu \in \mathfrak{M}[\omega, \varpi] \\ \infty & \text{otherwise.} \end{cases}$$

Define the function

$$\tilde{J} \colon \Theta \times \mathcal{M}(\mathcal{X} \times \mathcal{N}(\mathcal{X})) \to [0, \infty], \qquad \tilde{J}((\omega, \varpi), \nu) = \tilde{J}_{(\omega,\varpi)}(\nu).$$

Lemma 2.6.2. *$\tilde{J}$ is lower semicontinuous.*

Proof. Suppose $\theta_n := ((\omega_n, \varpi_n), \nu_n)$ converges to $\theta := ((\omega, \varpi), \nu)$ in $\Theta \times \mathcal{M}(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$. There is nothing to show if $\liminf_{\theta_n \to \theta} J(\theta_n) = \infty$. Otherwise, if (ϖ_n, ν_n) is sub-consistent for infinitely many n, then

$$\varpi(a,b) = \lim_{n\uparrow\infty} \varpi_n(a,b) \ge \liminf_{n\uparrow\infty} \langle \nu_n(\cdot, \ell), \ell(\cdot) \rangle (a,b) \ge \langle \nu(\cdot, \ell), \ell(\cdot) \rangle (a,b)\,,$$

hence (ϖ, ν) is sub-consistent. Similarly, if the first marginal of ν_n is ω_n, we see that the first marginal of ν is ω. We may therefore argue as in (2.5.7) to obtain

$$\begin{aligned} \liminf_{\theta_n \to \theta} J(\theta_n) = \liminf_{\theta_n \to \theta} H(\nu_n \,\|Q_n) \ge \lim_{\theta_n \to \theta} H(\nu_n \,\|\, Q_n) - H(\nu_n \,\|\, Q) \\ + \liminf_{\nu_n \to \nu} H(\nu_n \,\|Q) = H(\nu \,\|Q), \end{aligned}$$

where the last step is because of continuity of the entropy. This proves the lemma. ■

By Theorem 1.3.4 (see [Bi04]), the two previous lemmas and the large deviation principles we have established in Theorem 2.4.4 and 2.5.1 ensure that under $(\tilde{P}^n)$ the random variables $(\omega_n, \varpi_n, \nu_n)$ satisfy a large deviation principle on $\mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$ with good rate function

$$\hat{J}(\omega, \varpi, \nu) = \begin{cases} H(\omega \,\|\, \mu) + \frac{1}{2} \mathfrak{H}_C(\varpi \,\|\, \omega) + H(\nu \,\|Q)\,, & \text{if } \nu \in \mathfrak{M}[\omega, \varpi], \\ \infty\,, & \text{otherwise.} \end{cases}$$

By projection onto the last two components we obtain the large deviation principle as stated in Theorem 2.3.1 from the contraction principle, see e.g. [DZ98, Theorem 4.2.1].

2.7 Proof of the LDP for the Degree Distribution of Erdös-Renyi graph

In the case of an uncoloured Erdös-Renyi graph, the function C degenerates to a constant c, $L^2 = |E|/n \in [0,\infty)$ and $M = D \in \mathcal{M}(\mathbb{N} \cup \{0\})$. Theorem 2.3.1 and the contraction principle imply a large deviation principle for D with good rate function

$$\begin{aligned}\delta(d) &= \inf\big\{J(x,d)\colon x \geq 0\big\}\\ &= \inf\big\{H(d\,\|\,q_x) + \tfrac{1}{2}x\log x - \tfrac{1}{2}x\log c + \tfrac{1}{2}c - \tfrac{1}{2}x\colon \langle d\rangle \leq x\,\big\},\end{aligned}$$

where q_x is the Poisson distribution with parameter x. We denote by $\delta^x(d)$ the expression inside the infimum and consider the cases (i) $\langle d\rangle \geq c$ and (ii) $\langle d\rangle \leq c$ separately.

Case (i): For any $\varepsilon > 0$, we have

$$\begin{aligned}\delta^{\langle d\rangle+\varepsilon}(d) - \delta^{\langle d\rangle}(d) &= \tfrac{\varepsilon}{2} + \tfrac{\langle d\rangle-\varepsilon}{2}\log\tfrac{\langle d\rangle}{\langle d\rangle+\varepsilon} + \tfrac{\varepsilon}{2}\log\tfrac{\langle d\rangle}{c}\\ &\geq \tfrac{\varepsilon}{2} + \tfrac{\langle d\rangle-\varepsilon}{2}\left(\tfrac{-\varepsilon}{\langle d\rangle}\right) + \tfrac{\varepsilon}{2}\log\tfrac{\langle d\rangle}{c}\\ &> 0,\end{aligned}$$

so that the minimum is attained at $x = \langle d\rangle$.

Case (ii): Under our condition the equation $x = ce^{-2(1-\langle d\rangle/x)}$ has a unique solution, which satisfies $x \geq \langle d\rangle$. Elementary calculus shows that the global minimum of $y \mapsto \delta^y(d)$ on $(0,\infty)$ is attained at the value $y = x$, where x is the solution of our equation.

Chapter 3

The Asymptotic Equipartition Properties for Simple Hierarchical and Networked Structures

(*This material has also appeared in the preprint [DA06b].*)

3.1 Introduction

The underlying question is, how many bits are needed to store or transmit the information contained in a structured data consisting of n units connected by number of links?

Clearly, if no probabilistic structure is imposed, one needs of order n bits to transmit the units and of order n^2 bits to transmit the links of the network data structure. This is because 1 bit is needed to transmit each of the n unit and 1 bit is needed to transmit each of the $n(n-1)/2$ links in the structured data. By imposing a probabilistic structure which makes it very likely that the number of links is significantly smaller than maximum number $n(n-1)/2$, one can transmit the structure at much cheaper cost with arbitrarily high probability.

This is explained by the Shannon-McMillan-Breiman theorems for networked structures modelled as sparse coloured random graphs, and simple hierarchical structures modelled as multitype Galton-Watson trees. See, for example [CT91] for the classical Shannon-McMillan-Breiman theorem.

The rest of the chapter is divided into four main sections. Section 3.2 contains the AEP for hierarchically structured data. We state the main theorem, Theorem 3.2.1, in Subsection 3.2.1. The penultimate subsection of the section contains an application of the AEP to the data from mutation study. We end the discussion of the AEP for hierarchical structures by deriving the proof of Theorem 3.2.1. See Section 3.3.

We present in Section 3.4 the version of the AEP for simple networked structures. To be more specific, we state and interpret the main Theorems, Theorems 3.4.1, 3.4.2 in the first subsection, Subsection 3.4.1. In the next subsection, Subsection 3.4.2 we compute the asymptotic number of bits needed to code data from the metabolic network of the bacterium E.coli. The proof of the AEPs (for the network structured data) are carried out in Section 3.5.

Specifically, we derive WLLNs, see Lemmas 3.3.3 and 3.5.14, for the empirical offspring measure M_X of multitype Galton-Watson trees, and the empirical colour measure L^1 and the empirical pair measure L^2 of coloured random graphs from our LDP results. From Lemma 3.3.3 and the Perron-Frobenius Theorem, see, for example [DZ98, Theorem 3.1.1], we obtain Theorem 3.2.1, and from Lemma 3.5.14 we prove the other AEPs.

Definition 3.1.1. *We define the distribution $P_n\colon \mathcal{G}_n(\mathcal{X}) \to [0,1]$ of a coloured random graph X on n vertices by*

$$P_n(x) = \mathbb{P}\{\, X = x \,\}.$$

We write $\mathbb{P}_n\{X = x\} = \mathbb{P}\{\, X = x \,|\, |T| = n \,\}$ and define the distribution of multitype Galton-Watson tree x.

Definition 3.1.2. *The distribution of multitype Galton-Watson tree X with n vertices $P_n\colon \mathcal{T} \to [0,1]$ is defined by*

$$P_n(x) = \mathbb{P}_n\{\, X = x \,\}.$$

3.2 AEP for Hierarchical Structures

3.2.1 Main Theorem

We denote by $\mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$ the space of probability measures ν on $\mathcal{X} \times \mathcal{X}^*$ with $\int n\, \nu(da\,, dc) < \infty$, using the convention $c = (n, a_1, \ldots, a_n)$. We endow this space with the smallest topology which makes the functionals $\nu \mapsto \int f(b,c)\,\nu(db\,,dc)$ continuous, for $f : \mathcal{X} \times \mathcal{X}^* \to \mathbb{R}$ bounded. i.e. the weak topology.

Theorem 3.2.1. *Suppose $X = (X(v) : v \in V(T))$ is an irreducible, critical multitype Galton-Watson tree with finite type space $\mathcal{X}$ and bounded offspring kernel $\mathbb{Q}$. Then, for every $\varepsilon > 0$,*

$$\lim_{n\to\infty} \mathbb{P}_n\Big\{\Big| -\tfrac{1}{n}\log P_n(X) + \sum_{(a,c)\in\mathcal{X}\times\mathcal{X}^*} \pi(a)\mathbb{Q}\{c\,|\,a\}\log\mathbb{Q}\{c\,|\,a\}\Big| \geq \varepsilon\Big\} = 0.$$

We can extract from Theorem 3.2.1 the following useful information: To transmit the information contained in a large critical multitype Galton-Watson tree one needs, with probability close to 1, about

$$n\Big[-\frac{1}{\log 2}\sum_{(a,c)\in\mathcal{X}\times\mathcal{X}^*} \pi\otimes\mathbb{Q}(a,c)\log\mathbb{Q}\{c\,|\,a\}\Big] \qquad \text{bits,}$$

where n is the number of vertices in the tree.

3.2.2 Application to the Model of Mutation in Mitochondrial DNA

Mutations in mitochondrial DNA. Mitochondria are organelles in cells carrying their own DNA. Just like nuclear DNA, mtDNA is subject to mutations which may take the form of base substitutions, duplication or deletions. The population mtDNA is modelled by two-type process where the units are a (normal) and b (mutant), and the links are mother-child relations. A normal can give birth to either two normals or, if there is mutation, one normal and one mutant. Suppose the latter happens with probability or mutation rate α. Mutants can only give birth to mutants. A DNA molecule may also die without reproducing.

Let the survival probabilities be $p \in \big[0,\, \frac{1}{(2-\alpha)}\big]$ and $q \in \big[0,\, \frac{1}{2}\big]$ for normals and mutants resp. We assume that the population is started from one normal ancestor. Then the offspring kernel $\mathbb{Q}$ is given by $\mathbb{Q}\{(0,\emptyset)|a\} = 1-p$, $\mathbb{Q}\{(2,(a,b))|a\} = p\alpha$, $\mathbb{Q}\{(2,(a,a))|a\} = p(1-\alpha)$, $\mathbb{Q}\{(0,\emptyset)|b\} = 1-q$ and $\mathbb{Q}\{(2,(b,b))|b\} = q$. The process X is a multitype Galton-Watson process with matrix A (with index set

$\{a, b\}$) given by

$$A = \begin{pmatrix} p(2-\alpha) & 0 \\ p\alpha & 2q \end{pmatrix}.$$

Observe that this choice of p and q make the matrix A irreducible.

We restrict ourselves to the special case when $p = q = \frac{1}{2}$ and $\alpha > 0$. This case corresponds to the model for non-dividing tissue such as the brain. This means that the population of mtDNA is kept constant on average but that mitochondrial DNA keeps reproducing also in non-dividing cells. See, for example [OS02] and the references therein.

We observe that, in this special case X is critical and irreducible, with $\pi(a) = \pi(b) = \frac{1}{2}$. Therefore, by Theorem 3.2.1 one needs, with probability close to 1 approximately,

$$n\Big[1 - \tfrac{1}{\log 16}(\alpha \log \alpha + (1-\alpha)\log(1-\alpha))\Big] \text{ bits}, \qquad (3.2.1)$$

in order to store or transmit data from a model of non-dividing tissues.

For examples of data sources with tree structure, we refer to [KA02] or [Mo71].

3.3 Proof of the AEP for Hierarchical Structures

For our purpose we present a weak form of the large deviation principle for the empirical offspring measures in [DMS03].

To begin, we recall that for every multitype Galton-Watson tree X, the *empirical offspring measure* M_X is defined by

$$M_X(a,c) = \frac{1}{|T|} \sum_{v \in V} \delta_{(X(v), C(v))}(a,c), \quad \text{for } (a,c) \in \mathcal{X} \times \mathcal{X}^*.$$

Recall that ν *shift-invariant* if

$$\nu_1(a) = \sum_{(b,c) \in \mathcal{X} \times \mathcal{X}^*} m(a,c)\nu(b,c), \ \forall a \in \mathcal{X}.$$

3.3.1 LDPs for the Empirical Offspring Measures

To begin, we recall that for every multitype Galton-Watson tree X, the *empirical offspring measure* M_X is defined by

$$M_X(a,c) = \frac{1}{|T|} \sum_{v \in V} \delta_{(X(v),C(v))}(a,c), \ \text{ for } (a,c) \in \mathcal{X} \times \mathcal{X}^*.$$

Theorem 3.3.1 ([DMS03]). *Suppose that X is an irreducible, critical multitype Galton-Watson tree with bounded offspring kernel $\mathbb{Q}$, conditioned to have exactly n vertices. Then, for $n \to \infty$, the empirical offspring measure M_X satisfies a large deviation principle in $\mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$ with speed n and the convex, good rate function*

$$J(\nu) = \begin{cases} H(\nu \,\|\, \nu_1 \otimes \mathbb{Q}) & \text{if } \nu \text{ is shift-invariant,} \\ \infty & \text{otherwise.} \end{cases} \tag{3.3.1}$$

Alternatively, a new empirical measure we shall call the weighted empirical offspring measure S_X may be considered in the place of M_X. In the last part of this subsection, we define our new empirical offspring measure on $\mathcal{X} \times \mathcal{X}^*$ and state a version of Theorem 3.3.1. The proof follows from similar argument as that of Theorem 3.3.1 except that the measure M_X is replaced with S_X.

Definition 3.3.2. *Define for every multitype Galton-Watson tree X, the modified empirical offspring measure S_X by*

$$S_X(a,c) = \frac{1}{|E|} \sum_{v \in V} N(v)\delta_{(X(v),C(v))}(a,c), \ \textit{ for } (a,c) \in \mathcal{X} \times \mathcal{X}^*.$$

Next we state the LDP for the weighted empirical offspring measure S_X.

Theorem 3.3.3. *Suppose that X is an irreducible, critical multitype Galton-Watson tree with bounded offspring kernel $\mathbb{Q}$, conditioned to have exactly n vertices. Then, for $n \to \infty$, the weighted empirical offspring measure S_X satisfies a large deviation principle in $\mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$ with speed n and convex, good rate function,*

$$J(\nu) = \begin{cases} H\big((\frac{\nu}{n}) \,\|\, (\frac{\nu}{n})_1 \otimes \mathbb{Q}\big) & \text{if } (\frac{\nu}{n}) \text{ is shift-invariant,} \\ \infty & \text{otherwise.} \end{cases} \tag{3.3.2}$$

Remark 3 By contraction , see [DZ98, Theorem 4.2.1], one can recover from Theorem 3.3.3 the LDP for M_X in the weak topology.

Finally, we recall that the event $\{|T| = n\}$ under the law of an irreducible, critical multitype Galton-Watson tree T (under some moment condition) has probability which vanishes subexponentially at infinity. Denote by S the set of integers n where the probability $\mathbb{P}\{|T| = n\}$ is positive.

Lemma 3.3.4 ([DMS03]). *Suppose T is the random tree generated by an irreducible, critical multitype Galton-Watson tree with bounded offspring kernel. Then*

$$\lim_{\substack{n\to\infty \\ n\in S}} \frac{1}{n} \log \mathbb{P}\{|T| = n\} = 0.$$

Lemma 3.3.4 is key to the proof of the aforementioned AEP , the property for hierarchically structured data.

3.3.2 Open Problem

We observe that the exponential moments assumption of Theorem 3.3.1 is violated by geometric $\frac{1}{2}$ offspring distribution. i.e. $p(k) = 2^{-(k+1)}$, for $k \in \mathbb{N} \cup \{0\}$.

Consider Markov chain indexed by tree with offspring law $p(\cdot)$ defined as follows. First we sample a tree from probability measure $p(\cdot)$, and then, given the tree, we run a Markov chain on the vertices of the tree in such a way that the state of a vertex depends only on the state of the parent.

Question: Does the empirical offspring measure of the typed tree obey an LDP with explicit rate function under the joint law of tree and type ?

3.3.3 Weak Law of Large Numbers for the Empirical Offspring Measure.

Recall N_0 from the boundedness of $\mathbb{Q}$ and denote $\mathcal{X}_0^*$ by $\mathcal{X}_0^* = \bigcup_{n=0}^{N_0}\{n\} \times \mathcal{X}^n$. We equip it with the discrete topology. We recall also that π is the unique eigenvector (normalized to a probability vector) of the matrix A corresponding to the eigen value 1.

We derive from Theorem 3.3.1 the following weak law of large numbers.

Lemma 3.3.5. *Suppose that X is an irreducible, critical multitype Galton-Watson tree with bounded offspring law $\mathbb{Q}$, conditioned to have exactly n vertices. Then, for any $\varepsilon > 0$*

$$\lim_{n\to\infty} \mathbb{P}_n\Big\{ \max_{(a,c)\in\mathcal{X}\times\mathcal{X}_0^*} \big|M_X(a,c) - \pi\otimes\mathbb{Q}(a,c)\big| \geq \varepsilon\Big\} = 0. \tag{3.3.3}$$

Proof. Define the closed set

$$F = \big\{ \nu \in \mathcal{M}(\mathcal{X}\times\mathcal{X}_0) : \max_{(a,c)\in\mathcal{X}\times\mathcal{X}_0^*} \big|\nu(a,c) - \pi\otimes\mathbb{Q}(a,c)\big| \geq \varepsilon \,\big\}.$$

We observe from Theorem 3.3.1 that ,

$$\limsup_{n\to\infty} \tfrac{1}{n}\log\mathbb{P}_n\big\{ M_X \in F \,\big\} \leq -\inf_{\nu\in F} J(\nu). \tag{3.3.4}$$

We show by contradiction that the right hand side of (3.3.4) is negative.

To do this, we suppose that there exists sequence ν_n in F such that $J(\nu_n) \downarrow 0$. Then, because J is a good rate function and its level sets are compact, and by lower semicontinuouity of the mapping $\nu \mapsto J(\nu)$, there is a limit $\nu \in F$ with $J(\nu) = 0$. Hence, we have that ν is shift-invariant and $H(\nu \,\|\, \nu_1\otimes\mathbb{Q}) = 0$. This implies $\nu(a,c) = \nu_1\otimes\mathbb{Q}(a,c)$, for every $(a,c) \in \mathcal{X}\times\mathcal{X}^*$. Using shift-invariance of ν, for any $b \in \mathcal{X}$, we have

$$\sum_{(a,b)\in\mathcal{X}\times\mathcal{X}_0^*} \mathbb{Q}\{c\,|\,a\}m(b,c)\nu_1(a) = \sum_{(a,b)\in\mathcal{X}\times\mathcal{X}_0^*} \nu(a,c)m(b,c) = \nu_1(b).$$

This means that ν_1 is a non-negative eigenvector of A. By uniqueness of the Perron-Frobenius eigenvector, see, for example [DZ98, Theorem 3.1.1(d)], we infer that $\nu_1 = \pi$. This contradicts $\pi \otimes \mathbb{Q} \notin F$. ■

We now compute the distribution $P_n : \mathcal{T} \to [0,1]$ of X,

$$P_n(x) = \tfrac{\mu(x(\rho))}{\mathbb{P}\{|T|=n\}} \prod_{v \in V(T), |T|=n} \mathbb{Q}\{C(v) = c(v) \,|\, X(v) = x(v)\},$$

$(x(v), c(v))$ is the type, and the configuration of children of vertex v of $x \in \mathcal{T}$.

Therefore, we have that

$$-\tfrac{1}{n} \log P_n(x) = -\tfrac{1}{n} \log \mu(x(\rho)) + \tfrac{1}{n} \log \mathbb{P}\{|T| = n\} + \langle M_x, -\log \mathbb{Q}\rangle.$$

Now the term $\frac{1}{n}\log \mu(x(\rho))$ converges to zero, while the term $\frac{1}{n}\log \mathbb{P}\{|T| = n\}$ converges to zero because $\mathbb{Q}$ is bounded. See Lemma 3.3.4.

We observe that $-\log \mathbb{Q}$ is almost surely bounded on the support of M_X and therefore, by Lemma 3.3.5 we have

$$\langle M_x, -\log \mathbb{Q}\rangle \to \langle \pi \otimes \mathbb{Q}, -\log \mathbb{Q}\rangle,$$

which concludes the proof of Theorem 3.2.1.

3.4 AEPs for simple Networked Structures

3.4.1 Main Theorems

Theorem 3.4.1. *Suppose that X is a random coloured random graph with colour law $\mu\colon \mathcal{X} \to (0,1]$ and connection probabilities p_n such that $a_n^{-1} p_n(a,b) \to C(a,b)$, for some sequence (a_n) with $a_n \log n \to \infty$ and $\log a_n / \log n \to -1$. Then, for every $\varepsilon > 0$,*

$$\lim_{n\to\infty} \mathbb{P}\Big\{ \big| -\tfrac{1}{a_n^2 \log n} \log P_n(X) - \tfrac{1}{2} \sum_{a,b \in \mathcal{X}} \mu(a) C(a,b) \mu(b) \big| \ge \varepsilon \Big\} = 0.$$

In other words, in order to transmit a large random coloured random graph with

n vertices in the given regime one needs, with probability close to 1 about

$$\frac{a_n^2 \log n}{2\log 2} \sum_{a,b\in\mathcal{X}} \mu(a)C(a,b)\mu(b)$$

bits. The most interesting regime is when the cost of transmitting colours and transmitting edges is of comparable order, i.e. when

$$a_n = \tfrac{1}{n\log n}.$$

In this case one obtains the following Shannon-McMillan-Breiman theorem.

Theorem 3.4.2. *Suppose that X is a coloured random graph with colour law $\mu\colon \mathcal{X} \to (0,1]$ and connection probabilities p_n such that $(n\log n)\, p_n(a,b) \to C(a,b)$. Then, for every $\varepsilon > 0$,*

$$\lim_{n\to\infty} \mathbb{P}\Big\{ \big| -\tfrac{1}{n}\log P_n(X) - \tfrac{1}{2}\sum_{a,b\in\mathcal{X}} \mu(a)C(a,b)\mu(b) + \sum_{a\in\mathcal{X}} \mu(a)\log\mu(a) \big| \ge \varepsilon \Big\} = 0.$$

From Theorem 3.4.2 one can deduce that, the number of bits needed in order to code a networked data structure consisting of n units connected by number of order $n/\log n$ links is about nH, where H is the entropy defined by

$$H := \frac{1}{\log 2}\Big[\tfrac{1}{2}\sum_{a,b\in\mathcal{X}} \mu(a)C(a,b)\mu(b) - \sum_{a\in\mathcal{X}} \mu(a)\log\mu(a)\Big]. \tag{3.4.1}$$

3.4.2 Application to the Model of the Metabolic Network of the Bacterium E.Coli

We consider a metabolic network of the energy and biosynthesis metabolism of the bacterium E.coli. Here, the units represent substrates and products, and links represent interactions. Suppose half the nodes in the graph are of unit a (substrate) and half are of unit b (product), and link between pair of units (a,b) occur independently with connection probability $\frac{C(a,b)}{n}$, where $C : \{a,\, b\} \times \{a,\, b\} \to [0,\infty)$ is nonzero symmetric function and n the size of the graph. We write

$$H := \tfrac{1}{8\log 2}(2C(a,b) + C(a,a) + C(b,b)).$$

Then, by Theorem 3.4.1, one needs, with a probability close to 1 about,

$$(n\log n)\, H \ \text{ bits,}$$

to transmit the data contained in the metabolic network of the bacterium E.Coli.

3.5 Proofs of the AEPs for Networked Structures

3.5.1 Overview

Recall that the AEP is the analogue of the strong law of large numbers in information theory and that, it can be obtained from the WLLN for a carefully defined empirical measure. The WLLN for empirical measures on a finite space can often be deduced from an LDP for such a measure.

In this chapter we establish two main principles for the empirical colour measure L^1 and the empirical pair measure L^2 on sub- and supercritical coloured random graphs. All our proofs are based on the technique of change of measure. These principles are the main ingredients in the proof of the AEP for subcritical and supercritical networked structures.

To be more specific about our aim, we state all our principles in Subsection 3.5.2. The first principle is the joint large deviation principles for L^1 and L^2 defined on supercritical coloured random graphs. The second is LDP for the pair (L^1, L^2) of subs-critical random graphs. The subsequent sections contain the proofs of the principles. For supercritical coloured random graphs, we discuss the LDP on the scales n and $a_n n^2$ (where $a_n \to 0$) as the size of the graph goes to infinity in Subsections 3.5.3 and 3.5.4, respectively. Subsections 3.5.5 and 3.5.6 contain the proofs of the LDPs for (L^1, L^2) of subcritical coloured random graphs on the scales n and $a_n n^2$ (where $a_n \to 0$) in the weak topology. Finally, in Subsection 3.5.7 we prove the WLLNs for L^1 and L^2, and use them to derive our main theorems.

3.5.2 LDPs for the Empirical Colour Measure and the Empirical Pair Measure of Super-and Sub-critical Coloured Graphs

We recall that the empirical pair measure $L^2 \in \tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X})$ and the empirical colour measure $L^1 \in \mathcal{M}(\mathcal{X})$ are given by

$$L^2 = \frac{1}{a_n^2} \sum_{(u,v)\in E} [\delta_{(X(v),X(u))} + \delta_{(X(u),X(v))}],$$

$$L^1 = \frac{1}{n}\sum_{v\in V}\delta_{X(v)}.$$

Theorem 3.5.1 (Supercritical case). *Suppose that X is a random randomly coloured graph with colour law $\mu\colon \mathcal{X} \to (0,1]$ and connection probabilities $p_n : \mathcal{X}\times\mathcal{X} \to [0,1]$ satisfying $a_n^{-1}p_n(a,b) \to C(a,b)$, for some sequence (a_n) with $na_n \to \infty$ and $C : \mathcal{X}\times\mathcal{X} \to [0,\infty)$. Then, the pair (L^1, L^2) satisfies a large deviation principle as $n\to\infty$ in $\mathcal{M}(\mathcal{X})\times\tilde{\mathcal{M}}_*(\mathcal{X}\times\mathcal{X})$ with speed*

(i) *$a_n n^2$ and good rate function,*

$$I_1(\omega,\varpi) = \tfrac{1}{2}\mathfrak{H}_C(\varpi \,\|\, \omega). \tag{3.5.1}$$

(ii) *n and good rate function,*

$$I_2(\omega,\varpi) = \begin{cases} H(\omega \,\|\, \mu) & \text{if } \varpi = C\omega\otimes\omega, \\ \infty & \text{otherwise.} \end{cases} \tag{3.5.2}$$

Remark 4 Intuitively this means that, on the scale a_n^2 the colour law can be changed 'for free', whereas on the scale n once the colour law is fixed, the edge law has to be the typical one.

Theorem 3.5.2 (Sub-Critical). *Suppose that X is a random randomly coloured graph with colour law $\mu\colon \mathcal{X} \to (0,1]$ and connection probabilities $p_n : \mathcal{X}\times\mathcal{X} \to [0,1]$ satisfying $a_n^{-1}p_n(a,b) \to C(a,b)$, for some sequence (a_n) with $na_n \to 0$ and $C : \mathcal{X}\times\mathcal{X} \to [0,\infty)$. Then, the pair (L^1, L^2) satisfies a large deviation principle as $n\to\infty$ in $\mathcal{M}(\mathcal{X})\times\tilde{\mathcal{M}}_*(\mathcal{X}\times\mathcal{X})$ with speed*

(i) *$a_n n^2$ and good rate function,*

$$I_3(\omega,\varpi) = \begin{cases} \tfrac{1}{2}\mathfrak{H}_C(\varpi \,\|\, \omega) & \text{if } \omega = \mu, \\ \infty & \text{otherwise.} \end{cases} \tag{3.5.3}$$

(ii) *n and good rate function,*

$$I_4(\omega,\varpi) = H(\omega \,\|\, \mu). \tag{3.5.4}$$

Remark 5 Intuitively this means that, on the scale n the edge law can be changed 'for free', whereas on the scale a_n^2 the colour law cannot be changed.

3.5.3 Large Deviations for Super-Critical Coloured Random Graphs on the Scale n

Recall that $\mathcal{C}_2$ is the space of symmetric functions on $\mathcal{X} \times \mathcal{X}$ and $\mathcal{C}_1$ is the space of functions on $\mathcal{X}$. Define for $\tilde{f} \in \mathcal{C}_1$ $U_{\tilde{f}}$ by $U_{\tilde{f}} = \log \sum_{a \in \mathcal{X}} e^{\tilde{f}(a)} \mu(a)$.

We define for $\tilde{g} \in \mathcal{C}_2$ the symmetric function $\tilde{h}_n^{(2)} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ by

$$\tilde{h}_n^{(2)}(a,b) = \log \left[\left(1 - p_n(a,b) + p_n(a,b) e^{\tilde{g}(a,b)/na_n}\right)^{-n} \right]. \qquad (3.5.5)$$

Use $\tilde{f}$, $\tilde{g}$ to define (for sufficiently large n) a new random graph as follows:

- To the vertices $V = \{1, \ldots, n\}$ we assign colours from $\mathcal{X}$ independently and identically according to the colour law $\widetilde{\mu}$ defined by

$$\widetilde{\mu}(a) = e^{\tilde{f}(a) - U_{\tilde{f}}} \mu(a).$$

- Given any two vertices $u, v \in V$, with u carrying colour a and v carrying colour b connect vertex u to vertex v with probability

$$\tilde{p}_n(a,b) = \frac{p_n(a,b) e^{\tilde{g}(a,b)/na_n}}{1 - p_n(a,b) + p_n(a,b) e^{\tilde{g}(a,b)/na_n}}, \qquad (3.5.6)$$

  otherwise keep u and v disconnected.

For this new graph, observe $\widetilde{\mu}$ is a probability measure and further that

$$\tilde{p}_n(a,\, b) \in [0,\, 1], \; \forall a, b \in \mathcal{X}.$$

Denote by $\tilde{\mathbb{P}}$ the law of the new coloured random graph constructed from $\widetilde{\mu}$ and $\tilde{p}$.

We recall from Chapter 2 that $L^1_\Delta = \frac{1}{n} \sum_{u \in V} \delta_{(X(u), X(u))}$.

We note from the construction of the new graph that $\tilde{\mathbb{P}}$ is absolutely continuous with respect to $\mathbb{P}$, as for a coloured random graph X,

$$
\begin{aligned}
\frac{d\tilde{\mathbb{P}}}{d\mathbb{P}}(X) &= \prod_{u\in V} \tfrac{\tilde{\mu}(X(u))}{\mu(X(u))} \prod_{(u,v)\in E} \tfrac{\tilde{p}_n(X(u),X(v))}{p_n(X(u),X(v))} \prod_{(v,u)\notin E} \tfrac{1-\tilde{p}_n(X(u),X(v))}{1-p_n(X(u),X(v))} \\
&= \prod_{u\in V} \tfrac{\tilde{\mu}(X(u))}{\mu(X(u))} \prod_{(u,v)\in E} \tfrac{\tilde{p}_n(X(u),X(v))}{p_n(X(u),X(v))} \times \tfrac{n-np_n(X(u),X(v))}{n-n\tilde{p}_n(X(u),X(v))} \prod_{(u,v)\in \mathcal{E}} \tfrac{n-n\tilde{p}_n(X(u),X(v))}{n-np_n(X(u),X(v))} \\
&= \prod_{u\in V} e^{\tilde{f}(X(u))-U_{\tilde{f}}} \prod_{(u,v)\in E} e^{\tilde{g}(X(u),X(v))/na_n} \prod_{(u,v)\in \mathcal{E}} e^{\tilde{h}_n(2)(X(u),X(v))/n} \\
&= e^{n\,\langle L^1,\tilde{f}-U_{\tilde{f}}\rangle+n\,\langle \frac{1}{2}L^2,\tilde{g}\rangle+n\,\langle \frac{1}{2}L^1\otimes L^1,\tilde{h}_n^{(2)}\rangle-n\,\langle L^1_\Delta,\tilde{h}_n^{(2)}\rangle}.
\end{aligned} \tag{3.5.7}
$$

Upper Bound in Theorems 3.5.1(ii)

Define for $(\omega,\varpi)\in\mathcal{M}(\mathcal{X})\times\tilde{\mathcal{M}}_*(\mathcal{X}\times\mathcal{X})$, the function $\hat{I}_2(\omega,\,\varpi)$ by

$$
\hat{I}_2(\omega,\,\varpi)=\sup_{\substack{\tilde{f}\in\mathcal{C}_1\\ \tilde{g}\in\mathcal{C}_2}}\Big\{\sum_{a\in\mathcal{X}}\big(\tilde{f}(a)-U_{\tilde{f}}\big)\omega(a)+\sum_{a,b\in\mathcal{X}}\tfrac{1}{2}\tilde{g}(a,b)(\varpi(a,b)-C(a,b)\omega(a)\omega(b))\Big\}
\tag{3.5.8}
$$

Lemma 3.5.3. *For each closed set $F\subset\tilde{\mathcal{M}}_*(\mathcal{X}\times\mathcal{X})$,*

$$
\limsup_{n\to\infty}\tfrac{1}{n}\log\mathbb{P}\big\{(L^1,\,L^2)\in F\big\}\le-\inf_{(\omega,\,\varpi)\in F}\hat{I}_2(\omega,\,\varpi).
$$

Proof. Fix $\tilde{f}\in\mathcal{C}_1$. For any $\tilde{g}\in\mathcal{C}_2$, we define $\tilde{\beta}\colon\mathcal{X}\times\mathcal{X}\to\mathbb{R}$ by

$$
\tilde{\beta}(a,b)=-\tilde{g}(a,b)C(a,b).
$$

We notice from Lemma 2.4.8 that, $\lim_{n\to\infty}\tilde{h}_n^{(2)}(a,b)=\tilde{\beta}(a,b),\ \forall a,b\in\mathcal{X}$.

Using (3.5.7) we get

$$
e^{2\max_{a\in\mathcal{X}}|\tilde{\beta}(a,a)|}\ge\int e^{n\,\langle\frac{1}{2}L^1_\Delta,\tilde{h}_n^{(2)}\rangle}d\tilde{\mathbb{P}}=\mathbb{E}\Big\{e^{n\,\langle L^1,\tilde{f}-U_{\tilde{f}}\rangle+n\,\langle\frac{1}{2}L^2,\tilde{g}\rangle+n\,\langle\frac{1}{2}L^1\otimes L^1,\tilde{h}_n^{(2)}\rangle}\Big\}.
$$

Therefore, we have

$$
\limsup_{n\to\infty}\tfrac{1}{n}\log\mathbb{E}\Big\{e^{n\,\langle L^1,\tilde{f}-U_{\tilde{f}}\rangle+n\,\langle\frac{1}{2}\,L^2,\tilde{g}\rangle+n\,\langle\frac{1}{2}\,L^1\otimes L^1,\tilde{h}_n^{(2)}\rangle}\Big\}\le 0.
\tag{3.5.9}
$$

We now fix $\varepsilon>0$ and write $\hat{I}_2^\varepsilon(\omega,\,\varpi):=\min\{\hat{I}_2(\omega,\,\varpi),\varepsilon^{-1}\}-\varepsilon$.

Suppose $(\omega, \varpi) \in F$ and choose $\tilde{f} \in \mathcal{C}_1$, $\tilde{g} \in \mathcal{C}_2$ such that

$$\langle \omega, \tilde{f} - U_{\tilde{f}} \rangle + \tfrac{1}{2} \langle \varpi, \tilde{g} \rangle - \tfrac{1}{2} \langle \omega \otimes \omega, C\tilde{g} \rangle \geq \hat{I}_2^{\varepsilon}(\omega, \varpi).$$

Since $\mathcal{X}$ is finite, we can find open neighbourhoods B^2_{ϖ} and B^1_{ω} of ϖ, ω such that

$$\inf_{\tilde{\omega} \in B^1_{\omega},\, \tilde{\varpi} \in B^2_{\varpi}} \big\{ \langle \tilde{\omega}, \tilde{f} - U_{\tilde{f}} \rangle + \tfrac{1}{2} \langle \tilde{\varpi}, \tilde{g} \rangle - \tfrac{1}{2} \langle \tilde{\omega} \otimes \tilde{\omega}, C\tilde{g} \rangle \big\} \geq \hat{I}_2^{\varepsilon}(\omega, \varpi) - \varepsilon.$$

Using Chebysheff's inequality and (3.5.9), we have that

$$\begin{aligned}
&\limsup_{n\to\infty} \tfrac{1}{n} \log \mathbb{P}\big\{ (L^1, L^2) \in B^1_{\omega} \times B^2_{\varpi} \big\} \\
&\quad \leq \limsup_{n\to\infty} \tfrac{1}{n} \log \mathbb{E}\Big\{ e^{n \langle L^1, \tilde{f} - U_{\tilde{f}} \rangle + n \langle \frac{1}{2} L^2, \tilde{g} \rangle + n \langle \frac{1}{2} L^1 \otimes L^1, \tilde{h}_n^{(2)} \rangle} \Big\} - \hat{I}_2^{\varepsilon}(\omega, \varpi) + \varepsilon \\
&\quad \leq -\hat{I}_2^{\varepsilon}(\omega, \varpi) + \varepsilon.
\end{aligned} \tag{3.5.10}$$

Use Lemma 2.4.9 with $\alpha = \varepsilon^{-1}$ to choose $N(\varepsilon) \in \mathbb{N}$ such that

$$\limsup_{n\to\infty} \tfrac{1}{n} \log \mathbb{P}\Big\{ |E| > a_n n^2 N(\varepsilon) \Big\} = -\infty.$$

For this N define the set $K_{N(\varepsilon)}$ by

$$K_{N(\varepsilon)} = \big\{ (\omega, \varpi) \in \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X}) : \|\varpi\| \leq 2N(\varepsilon) \big\}.$$

The set $K_{N(\varepsilon)} \cap F$ is compact and therefore may be covered by finitely many sets

$$B^1_{\omega_r} \times B^2_{\varpi_r}, r = 1, \ldots, m \text{ with } (\omega_r, \varpi_r) \in F \text{ for } r = 1, \ldots, m.$$

Hence, we have

$$\mathbb{P}\big\{ (L^1, L^2) \in F \big\} \leq \sum_{r=1}^{m} \mathbb{P}\big\{ (L^1, L^2) \in B^1_{\omega_r} \times B^2_{\varpi_r} \big\} + \mathbb{P}\big\{ (L^1, L^2) \notin K_{N(\varepsilon)} \big\}.$$

We may now use (3.5.10) to obtain, for all sufficiently small $\varepsilon > 0$,

$$\begin{aligned}
\limsup_{n\to\infty} \tfrac{1}{n} \log \mathbb{P}\big\{ (L^1, L^2) \in F \big\} &\leq \max_{r=1}^{m} \limsup_{n\to\infty} \tfrac{1}{n} \log \mathbb{P}\big\{ (L^1, L^2) \in B^1_{\omega_r} \times B^2_{\varpi_r} \big\} \vee -\varepsilon^{-1} \\
&\leq - \inf_{(\omega, \varpi) \in F} \hat{I}_2^{\varepsilon}(\omega, \varpi) \vee -\varepsilon^{-1} + \varepsilon.
\end{aligned}$$

Taking $\varepsilon \downarrow 0$ we have the desired statement. ∎

We solve the variational problem on the right side of equation (3.5.8).

Lemma 3.5.4. *$\hat{I}_2(\omega,\,\varpi) = H(\omega\,\|\,\mu)$ if (and only if) $\varpi = C\omega\otimes\omega$, and ∞ otherwise.*

Proof. Suppose that $\varpi \neq C\omega\otimes\omega$. Then there exists $a_0, b_0 \in \mathcal{X}$ such

$$\varpi(a_0,b_0) > C(a_0,b_0)\omega(a_0)\otimes\omega(b_0).$$

Define for this $a_0, b_0 \in \mathcal{X}$ the symmetric function $\tilde{g}$ by

$$\tilde{g}(a,b) = K(\mathbb{1}_{(a_0,b_0)}(a,b) + \mathbb{1}_{(b_0,a_0)})(a,b), \tag{3.5.11}$$

for $a, b \in \mathcal{X}$ and $K > 0$.

Considering this $\tilde{g}$ in (3.5.8) we have

$$\sum_{a,b\in\mathcal{X}} \tfrac{1}{2}\tilde{g}(a,b)\varpi(a,b) + \sum_{a,b\in\mathcal{X}} -\tfrac{1}{2}\tilde{g}(a,b)C(a,b)\omega(a)\omega(b)$$
$$= K(\varpi(a_0,b_0) - C(a_0,b_0)\omega(a_0)\omega(b)) \xrightarrow{K\uparrow\infty} \infty.$$

Suppose that $\varpi = C\omega\otimes\omega$. Then, by the variational characterization of the relative entropy we have

$$I_2(\omega,\,\varpi) = H(\omega\,\|\,\mu),$$

which ends the proof of the upper bounds. ∎

Lower Bounds Theorems 3.5.1(ii)

Lemma 3.5.5. *For every open set $O \subset \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}_*(\mathcal{X}\times\mathcal{X})$,*

$$\liminf_{n\to\infty} \tfrac{1}{n}\log\mathbb{P}\big\{(L^1,L^2)\in O\big\} \geq -\inf_{(\omega,\,\varpi)\in O} I_2(\omega,\,\varpi). \tag{3.5.12}$$

Proof. Suppose $(\omega,\,\varpi)\in O$ is such that we have $\varpi = C\omega\otimes\omega$.

Set $\tilde{g}(a,b) = 0$, $\forall a,b\in\mathcal{X}$ and define $\tilde{f}_\omega\colon \mathcal{X}\to\mathbb{R}$ by

$$\tilde{f}_\omega(a) = \begin{cases} \log\frac{\omega(a)}{\mu(a)}, & \text{if } \omega(a) > 0,\\ 0, & \text{otherwise.}\end{cases}$$

Note by this choice of $\tilde{g}$ we have $\lim_{n\to\infty}\tilde{h}_n^{(2)}(a,b) = 0$, $\forall a,b\in\mathcal{X}$.

Choose B^1_ω, B^2_ϖ open neighbourhoods of ω, ϖ, such that $B^1_\omega \times B^2_\varpi \subset O$ and

$$\forall(\tilde\omega, \tilde\varpi) \in B^1_\omega \times B^2_\varpi, \qquad \text{we have} \qquad \langle \tilde f_\omega, \omega\rangle - \varepsilon \le \langle \tilde f_\omega, \tilde\omega\rangle.$$

We use the probability measure $\tilde{\mathbb{P}}$ given by $\tilde g_\varpi$. We observe that the colour law is ω and the connection probabilities satisfy

$$a_n^{-1}\tilde p_n(a,b) \xrightarrow{n\uparrow\infty} \tilde C(a,b) := \tfrac{\varpi(a,b)}{\omega(a)\omega(b)}.$$

Therefore, using (3.5.7) we have that

$$\begin{aligned}\mathbb{P}\Big\{(L^1, L^2) \in O\Big\} \ge \tilde{\mathbb{E}}\Big\{\tfrac{d\mathbb{P}}{d\tilde{\mathbb{P}}}(X)\mathbb{1}_{\{(L^1,L^2)\in B^1_\omega\times B^2_\varpi\}}\Big\} &= \tilde{\mathbb{E}}\Big\{e^{-n\langle L^1, \tilde f_\omega\rangle}\mathbb{1}_{\{(L^1,\,L^2)\in B^1_\omega\times B^2_\varpi\}}\Big\}\\ &\ge e^{-n\langle\omega,\,\tilde f_\omega\rangle - n\varepsilon} \times \tilde{\mathbb{P}}\Big\{(L^1, L^2) \in B^1_\omega \times B^2_\varpi\Big\}.\end{aligned}$$

Therefore, we have

$$\begin{aligned}\liminf_{n\to\infty} \tfrac{1}{n}\log\mathbb{P}\{(L^1, L^2) \in O\} \ge &-\langle \omega,\, \tilde f_\omega\rangle - \varepsilon\\ &+ \liminf_{n\to\infty}\tfrac{1}{n}\log\tilde{\mathbb{P}}\{(L^1, L^2) \in B^1_\omega \times B^2_\varpi\}.\end{aligned}$$

The result follows once we prove that

$$\liminf_{n\to\infty}\tfrac{1}{n}\log\tilde{\mathbb{P}}\{(L^1, L^2) \in B^1_\omega \times B^2_\varpi\} = 0. \tag{3.5.13}$$

We use the upper bound (but now with the law $\mathbb{P}$ replaced by $\tilde{\mathbb{P}}$) to prove (3.5.13).

Therefore, we have that

$$\limsup_{n\to\infty}\tfrac{1}{n}\log\tilde{\mathbb{P}}\{(L^1, L^2) \in (B^1_\omega \times B^2_\varpi)^c\} \le -\inf_{(\tilde\omega,\tilde\varpi)\in\tilde F}\tilde I_2(\tilde\omega, \tilde\varpi),$$

$$\tilde I_2(\tilde\omega, \tilde\varpi) = \begin{cases} H(\tilde\omega\,\|\,\omega) & \text{if } \tilde\varpi = \tilde C\tilde\omega\otimes\tilde\omega,\\ \infty & \text{otherwise,}\end{cases}$$

where $\tilde F = (B^1_\omega \times B^2_\varpi)^c$.

It therefore suffices to show that the infimum is positive.

Suppose for contradiction that there exists a sequence $(\tilde{\omega}_n, \tilde{\varpi}_n) \in \tilde{F}$ with $\tilde{I}_2(\tilde{\omega}_n, \tilde{\varpi}_n) \downarrow 0$. Then, since $\tilde{I}_2$ is a good rate function and its level sets are compact, and the mapping $(\tilde{\omega}, \tilde{\varpi}) \mapsto \tilde{I}_2(\tilde{\omega}, \tilde{\varpi})$ is lower semicontinuous, we can construct a limit point $(\tilde{\omega}, \tilde{\varpi}) \in \tilde{F}$ with $\tilde{I}_2(\tilde{\omega}, \tilde{\varpi}) = 0$. By Lemma 3.5.4 this implies $H(\tilde{\omega} \,\|\, \omega) = 0$ and $\tilde{\varpi} = C\tilde{\omega} \otimes \tilde{\omega}$, hence $\tilde{\omega} = \omega$, and $\tilde{\varpi} = \tilde{C}\tilde{\omega} \otimes \tilde{\omega} = \varpi$. This contradicts $(\tilde{\omega}, \tilde{\varpi}) \in \tilde{F}$.

■

3.5.4 Large Deviations for Super-Critical Coloured Random Graphs on the Scale $a_n n^2$

We recall that for $\tilde{f} \in \mathcal{C}_1$ the constant $U_{\tilde{f}}$ is given by

$$U_{\tilde{f}} = \log \sum_{a \in \mathcal{X}} e^{\tilde{f}(a)} \mu(a).$$

Further, we define for $\tilde{g} \in \mathcal{C}_2$ the function $\tilde{h}_n^{(1)} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is given by

$$\tilde{h}_n^{(1)}(a,b) = -\log \left[\left(1 - p_n(a,b) + p_n(a,b) e^{\tilde{g}(a,b)} \right)^{1/a_n} \right].$$

Define for $\tilde{f}$, $\tilde{g}$ (for sufficiently large n) a new random graph in the following way:

- Assign to the n vertices in V colours from $\mathcal{X}$ independently and identically according to the colour law $\widetilde{\mu}$ defined by

$$\widetilde{\mu}(a) = e^{\tilde{f}(a) - U_{\tilde{f}}} \mu(a). \tag{3.5.14}$$

- Given any two vertices $u, v \in V$, with u carrying colour a and v carrying colour b connect vertex u to vertex v with probability

$$\tilde{p}_n(a,b) = \tfrac{p_n(a,b) e^{\tilde{g}(a,b)}}{1 - p_n(a,b) + p_n(a,b) e^{\tilde{g}(a,b)}}, \tag{3.5.15}$$

otherwise keep u and v disconnected.

Note the colour law $\tilde{\mu}$ is a probability measure and the connection probabilities

$$\tilde{p}_n(a,\, b) \in [0,\, 1],\ \forall a, b \in \mathcal{X}.$$

We denote by $\tilde{\mathbb{P}}$ the law of the coloured random graph obtained from $\widetilde{\mu}$ and $\tilde{p}_n$.

Write

$$L^2_\Delta := \tfrac{1}{n^2}\sum_{u\in V}\delta_{(X(u),X(u))}.$$

By construction $\tilde{\mathbb{P}}$ is absolutely continuous with respect to $\mathbb{P}$, as for a coloured random graph X,

$$\begin{aligned}
\frac{d\tilde{\mathbb{P}}}{d\mathbb{P}}(X) &= \prod_{u\in V}\tfrac{\tilde{\mu}(X(u))}{\mu(X(u))}\prod_{(u,v)\in E}\tfrac{\tilde{p}_n(X(u),X(v))}{p_n(X(u),X(v))}\prod_{(v,u)\notin E}\tfrac{1-\tilde{p}_n(X(u),X(v))}{1-p_n(X(u),X(v))}\\
&= \prod_{u\in V}e^{\tilde{f}(X(u))-U_{\tilde{f}}}\prod_{(u,v)\in E}\tfrac{\tilde{p}_n(X(u),X(v))}{p_n(X(u),X(v))}\times\tfrac{n-n\tilde{p}_n(X(u),X(v))}{n-np_n(X(u),X(v))}\prod_{(u,v)\in\mathcal{E}}\tfrac{n-np_n(X(u),X(v))}{n-n\tilde{p}_n(X(u),X(v))}\\
&= \prod_{u\in V}e^{\tilde{f}(X(u))-U_{\tilde{f}}}\prod_{(u,v)\in E}e^{\tilde{g}(X(u),\,X(v))}\prod_{(u,v)\in\mathcal{E}}e^{a_n\tilde{h}_n^{(1)}(X(u),\,X(v))}\\
&= e^{n\,\langle L^1,\,\tilde{f}-U_{\tilde{f}}\rangle+a_nn^2\,\langle\frac{1}{2}L^2,\,\tilde{g}\rangle+a_nn^2\,\langle\frac{1}{2}L^1\otimes L^1,\,\tilde{h}_n^{(1)}\rangle-a_nn^2\,\langle\frac{1}{2}L^2_\Delta,\,\tilde{h}_n^{(1)}\rangle}.
\end{aligned}\qquad(3.5.16)$$

Upper Bound in Theorem 3.5.1(i)

To begin we obtain the upper bound in a variational formulation.

We define for $(\omega,\varpi)\in\mathcal{M}(\mathcal{X})\times\tilde{\mathcal{M}}_*(\mathcal{X}\times\mathcal{X})$ the rate function $\hat{I}_1$ by

$$\hat{I}_1(\omega,\varpi)=\sup_{\tilde{g}\in\mathcal{C}_2}\Big\{\sum_{a,b\in\mathcal{X}}\tfrac{1}{2}\tilde{g}(a,b)\varpi(a,b)+\sum_{a,b\in\mathcal{X}}\tfrac{1}{2}(1-e^{\tilde{g}(a,b)})C(a,b)\omega(a)\omega(b)\Big\}.\qquad(3.5.17)$$

Lemma 3.5.6. *For each closed set $F\subset\mathcal{M}(\mathcal{X})\times\tilde{\mathcal{M}}_*(\mathcal{X}\times\mathcal{X})$,*

$$\limsup_{n\to\infty}\tfrac{1}{a_nn^2}\log\mathbb{P}\big\{(L^1,\,L^2)\in F\big\}\le-\inf_{(\omega,\,\varpi)\in F}\hat{I}_1(\omega,\,\varpi).$$

Proof. For any $\tilde{g}\in\mathcal{C}_2$ we define $\tilde{\beta}\colon\mathcal{X}\times\mathcal{X}\to\mathbb{R}$ by

$$\tilde{\beta}(a,b)=(1-e^{\tilde{g}(a,b)})C(a,b).$$

From Lemma 2.4.8 we note that we have $\lim_{n\to\infty}\tilde{h}_n^{(1)}(a,b)=\tilde{\beta}(a,b)$, $\forall a,b\in\mathcal{X}$.

We take $\tilde{f}(a)=0$, $\forall a\in\mathcal{X}$ and use (3.5.16) to obtain

$$e^{2na_n\max_{a\in\mathcal{X}}|\tilde{\beta}(a,a)|}\geq\iint e^{a_nn^2\langle\frac{1}{2}L^2_{\Delta},\tilde{h}_n^{(1)}\rangle}d\tilde{\mathbb{P}}=\mathbb{E}\Big\{e^{a_nn^2\langle\frac{1}{2}L^2,\tilde{g}\rangle+a_nn^2\langle\frac{1}{2}L^1\otimes L^1,\tilde{h}_n^{(1)}\rangle}\Big\}.$$

Therefore,

$$\limsup_{n\to\infty}\tfrac{1}{a_nn^2}\log\mathbb{E}\Big\{e^{a_nn^2\langle\frac{1}{2}L^2,\tilde{g}\rangle+a_nn^2\langle\frac{1}{2}L^1\otimes L^1,\tilde{h}_n^{(1)}\rangle}\Big\}\leq 0. \tag{3.5.18}$$

Fix $\varepsilon>0$ and take $\hat{I}_1^{\varepsilon}(\omega,\,\varpi)=\min\{\hat{I}_1(\omega,\,\varpi),\varepsilon^{-1}\}-\varepsilon$. Suppose $(\omega,\,\varpi)\in F$.

Choose $\tilde{g}\in\mathcal{C}_2$ such that $\frac{1}{2}\langle\varpi,\,\tilde{g}\rangle+\frac{1}{2}\langle\omega\otimes\omega,\,\tilde{\beta}\rangle\geq\hat{I}_1^{\varepsilon}(\omega,\varpi)$.

By finiteness of $\mathcal{X}$, we can find open neighbourhoods B^1_{ω}, B^2_{ϖ} of ω, ϖ such that

$$\inf_{\tilde{\omega}\in B^1_{\omega},\,\tilde{\varpi}\in B^2_{\varpi}}\big\{\tfrac{1}{2}\langle\tilde{\varpi},\,\tilde{g}\rangle+\tfrac{1}{2}\langle\tilde{\omega}\otimes\tilde{\omega},\,\tilde{\beta}\rangle\big\}\geq\hat{I}_1^{\varepsilon}(\omega,\,\varpi)-\varepsilon.$$

By Chebysheff's inequality and (3.5.18), we have that

$$\begin{aligned}
&\limsup_{n\to\infty}\tfrac{1}{a_nn^2}\log\mathbb{P}\big\{(L^1,\,L^2)\in B^1_{\omega}\times B^2_{\varpi}\big\}\\
&\leq\limsup_{n\to\infty}\tfrac{1}{a_nn^2}\log\mathbb{E}\Big\{e^{a_nn^2\langle\frac{1}{2}L^2,\tilde{g}\rangle+a_nn^2\langle\frac{1}{2}L^1\otimes L^1,\tilde{h}_n^{(1)}\rangle}\Big\}-\hat{I}_1^{\varepsilon}(\omega,\,\varpi)+\varepsilon\\
&\qquad\leq-\hat{I}_1^{\varepsilon}(\omega,\,\varpi)+\varepsilon.
\end{aligned}\tag{3.5.19}$$

We use Lemma (2.4.9) to choose $N(\varepsilon)\in\mathbb{N}$ (with $\alpha=\varepsilon^{-1}$) such that

$$\limsup_{n\to\infty}\tfrac{1}{a_nn^2}\log\mathbb{P}\Big\{|E|>a_nn^2N(\varepsilon)\Big\}\leq-\varepsilon^{-1}.$$

Define for this N, the set $K_{N(\varepsilon)}$ by

$$K_{N(\varepsilon)}=\big\{(\omega,\,\varpi)\in\mathcal{M}(\mathcal{X})\times\tilde{\mathcal{M}}_*(\mathcal{X}\times\mathcal{X}):\|\varpi\|\leq 2N(\varepsilon)\big\}.$$

Now, observe that $K_{N(\varepsilon)}\cap F$ is compact and therefore may be covered by finitely many sets

$$B^1_{\omega}\times B^2_{\varpi_r},r=1,...,m\text{ with }\varpi_r\in F\text{ for }r=1,...,m.$$

Hence, we have that

$$\mathbb{P}\big\{(L^1,\,L^2)\in F\big\}\leq\sum_{r=1}^{m}\mathbb{P}\big\{(L^1,\,L^2)\in B^1_{\omega_r}\times B^2_{\varpi_r}\big\}+\mathbb{P}\big\{(L^1,\,L^2)\notin K_{N(\varepsilon)}\big\}.$$

Using (3.5.19) for small enough $\varepsilon > 0$, we obtain

$$\begin{aligned}\limsup_{n\to\infty} \tfrac{1}{a_n n^2} \log \mathbb{P}\big\{(L^1, L^2) \in F\big\} \\ &\le \max_{r=1}^{m} \limsup_{n\to\infty} \tfrac{1}{a_n n^2} \log \mathbb{P}\big\{(L^1,\, L^2) \in B^1_{\omega_r} \times B^2_{\varpi_r}\big\} \vee -\varepsilon^{-1} \\ &\le -\hat{I}_1^{\varepsilon}(\omega,\, \varpi) \vee -\varepsilon^{-1} + \varepsilon.\end{aligned}$$

Taking $\varepsilon \downarrow 0$ we have the desired statement.

■

We identify the rate function by solving the variational problem in the right side of equation (3.5.17).

Lemma 3.5.7. *For any* $(\omega,\, \varpi) \in \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X})$ *we have*

(i) $\hat{I}_1(\omega,\, \varpi) = I_1(\omega,\, \varpi)$ *and (ii)* I_1 *is good rate function.*

Proof. (i) Suppose $\varpi \not\ll C\omega \otimes \omega$. Then there exists $a_0,\, b_0 \in \mathcal{X}$ with

$$C(a_0, b_0)\omega(a_0)\omega(b_0) = 0 \qquad \text{and} \qquad \varpi(a_0, b_0) > 0.$$

For this (a_0, b_0) we define the symmetric function $\tilde{g}$ by

$$\tilde{g}(a, b) = \log(K(\mathbb{1}_{(a_0,b_0)}(a, b) + \mathbb{1}_{(b_0,a_0)}(a, b)) + 1),$$

for $a, b \in \mathcal{X}$ and $K > 0$.

Considering our $\tilde{g}$ in (3.5.17) we have

$$\begin{aligned}\sum_{a,b\in\mathcal{X}} \tfrac{1}{2}\tilde{g}(a, b)\varpi(a, b) + \sum_{a,b\in\mathcal{X}} \tfrac{1}{2}(1 - e^{\tilde{g}(a,b)})C(a, b)\omega(a)\omega(b) \\ = \log(K + 1)(\varpi(a_0, b_0)) \xrightarrow{K\uparrow\infty} \infty.\end{aligned}$$

Suppose that $\varpi \ll C\omega \otimes \omega$. Then, we have

$$\begin{aligned}\hat{I}_1(\omega, \varpi) = \tfrac{1}{2} \sup_{g\in\mathcal{C}_2} \Big\{ \sum_{a,b\in\mathcal{X}} g(a, b)\varpi(a, b) - \sum_{a,b\in\mathcal{X}} e^{g(a,b)}C(a, b)\omega(a)\omega(b)\Big\} \\ + \tfrac{1}{2} \sum_{a,b\in\mathcal{X}} C(a, b)\omega(a)\omega(b).\end{aligned} \tag{3.5.20}$$

Using the substitution $h = e^g \frac{C\omega\otimes\omega}{\varpi}$ and $\sup_{x>0} \log x - x = -1$ we obtain the expression

$$\sup_{g\in\mathcal{C}_2}\Big\{\sum_{a,b\in\mathcal{X}} g(a,b)\varpi(a,b) - \sum_{a,b\in\mathcal{X}} e^{g(a,b)}C(a,b)\omega(a)\omega(b)\Big\}$$

$$\sup_{\substack{h\in\mathcal{C}_2\\ h\geq 0}}\sum_{a,b\in\mathcal{X}}\Big[\log\Big(h(a,b)\frac{\varpi(a,b)}{C(a,b)\omega(a)\omega(b)}\Big) - h(a,b)\Big]\varpi(a,b)$$

$$= \sup_{\substack{h\in\mathcal{C}_2\\ h\geq 0}}\sum_{a,b\in\mathcal{X}}\big(\log h(a,b) - h(a,b)\big)\varpi(a,b) + \sum_{a,b\in\mathcal{X}}\log\Big(\frac{\varpi(a,b)}{C(a,b)\omega(a)\omega(b)}\Big)\varpi(a,b)$$

$$= -\|\varpi\| + H(\varpi\,\|\,C\omega\otimes\omega).$$

This yields that

$$\hat{I}_1(\omega,\,\varpi) = I_1(\omega,\,\varpi).$$

(ii) This follows from the proof of Lemma 2.4.13. Recall from the proof of Lemma 2.4.13 that $\mathfrak{H}_C(\varpi\,\|\,\omega) = \frac{1}{2}H\big(\varpi\,\|\,C\omega\otimes\omega\big) + \frac{1}{2}\|C\omega\otimes\omega\| - \frac{1}{2}\|\varpi\|$ is a good rate function and that for all $\alpha < \infty$, its level sets are the bounded, closed set $\{(\omega,\varpi)\in\mathcal{M}(\mathcal{X})\times\tilde{\mathcal{M}}(\mathcal{X}\times\mathcal{X})\colon \mathfrak{H}_C(\varpi\,\|\,\omega)\leq\alpha\}$ and so are compact. This implies I_1 is good rate function.

■

Lower Bound in Theorem 3.5.1(i). We use the upper bound (but now with the law $\mathbb{P}$ replaced by $\tilde{\mathbb{P}}$) to establish the lower bound for some open set $O\subset\mathcal{M}(\mathcal{X})\times\tilde{\mathcal{M}}_*(\mathcal{X}\times\mathcal{X})$.

Lemma 3.5.8. *For every open set $O\subset\mathcal{M}(\mathcal{X})\times\tilde{\mathcal{M}}_*(\mathcal{X}\times\mathcal{X})$.*

$$\liminf_{n\to\infty}\frac{1}{a_n n^2}\log\mathbb{P}\big\{(L^1,\,L^2)\in O\big\} \geq -\inf_{(\tilde{\omega},\,\tilde{\varpi})\in O} I_1(\omega,\,\varpi). \tag{3.5.21}$$

Proof. Suppose $(\omega,\,\varpi)\in O$ with $\varpi \ll C\omega\otimes\omega$. We define the function $\tilde{f}_\omega\colon\mathcal{X}\to\mathbb{R}$ by

$$\tilde{f}_\omega(a) = \begin{cases}\log\frac{\omega(a)}{\mu(a)}, & \text{if } \omega(a)>0,\\ 0, & \text{otherwise,}\end{cases}$$

and the symmetric function $\tilde{g}_\varpi\colon\mathcal{X}\times\mathcal{X}\to\mathbb{R}$ by

$$\tilde{g}_\varpi(a,b) = \begin{cases}\log\frac{\varpi(a,b)}{C(a,b)\omega(a)\omega(b)}, & \text{if } \varpi(a,b)>0,\\ 0, & \text{otherwise.}\end{cases}$$

We recall that $\tilde{h}_n^{(1)}(a,b) = -\log\left[1 - p_n(a,b) + p_n(a,b)e^{\tilde{g}_\varpi(a,b)}\right]^{1/a_n}$, for $a, b \in \mathcal{X}$.

Define the symmetric function $\tilde{\beta}_\varpi(a,b)$ by

$$\tilde{\beta}_\varpi(a,b) := \lim_{n\to\infty} \tilde{h}_n^{(n)}(a,b) = C(a,b)(1 - e^{g_\varpi(a,b)}).$$

Choose B_ω^1, B_ϖ^2 open neighbourhoods of ω, ϖ such that $B_\omega^1 \times B_\varpi^2 \subset O$ and

$$\begin{gathered}\forall(\tilde{\omega}, \tilde{\varpi}) \in B_\omega^1 \times B_\varpi^2, \\ \langle \varpi,\, \tilde{g}_\varpi\rangle + \langle \omega\otimes\omega,\, \tilde{\beta}_\varpi\rangle - \varepsilon \le \langle \tilde{\varpi},\, \tilde{g}_\varpi\rangle + \langle \tilde{\omega}\otimes\tilde{\omega},\, \tilde{\beta}_\varpi\rangle.\end{gathered}$$

We note that, the coloured random graph obtained from the function $\tilde{g}_\varpi$ has colour law ω and connection probabilities satisfying

$$a_n^{-1}\tilde{p}_n(a,b) \xrightarrow{n\uparrow\infty} \tilde{C}(a,b) := \tfrac{\varpi(a,b)}{\omega(a)\omega(b)}.$$

Write $m := 0 \wedge \min\limits_{a\in\mathcal{X}} \tilde{\beta}_\varpi(a,a)$, and $l := 0 \wedge \min\limits_{a\in\mathcal{X}} \tilde{f}(a)$. Hence, by (3.5.16) we have

$$\begin{aligned}&\mathbb{P}\Big\{(L^1,\, L^2) \in O\Big\} \ge \tilde{\mathbb{E}}\Big\{\tfrac{d\mathbb{P}}{d\tilde{\mathbb{P}}}(X)\mathbb{1}_{\{(L^1,\, L^2)\in B_\omega^1\times B_\varpi^2\}}\Big\} \\ &\quad = \tilde{\mathbb{E}}\Big\{e^{-n\langle L^1,\, \tilde{f}_\omega\rangle - a_n n^2\,\langle \frac{1}{2}\, L^2,\, \tilde{g}_\varpi\rangle - a_n n^2\,\langle \frac{1}{2}\, L^1\otimes L^1,\, \tilde{h}_n^{(1)}\rangle + n^2 a_n\langle \frac{1}{2}L_\Delta^2,\, \tilde{h}_n^{(1)}\rangle} \times \mathbb{1}_{\{(L^1,\, L^2)\in B_\omega^1\times B_\varpi^2\}}\Big\} \\ &\quad \ge e^{-nl - a_n n^2\,\langle\varpi,\, \tilde{g}_\varpi\rangle/2 - a_n n^2\,\langle\, \omega\otimes\omega,\, \tilde{\beta}\rangle/2 + a_n m/4 - a_n n^2\,\varepsilon/2} \times \tilde{\mathbb{P}}\Big\{(L^1,\, L^2) \in B_\omega^1 \times B_\varpi^2\Big\}.\end{aligned}$$

Therefore, we have

$$\begin{aligned}\liminf_{n\to\infty} \tfrac{1}{a_n n^2} \log \mathbb{P}\big\{(L^1,\, L^2) \in O\big\} &\ge -\tfrac{1}{2}\,\langle \tilde{g}, \varpi\rangle - \tfrac{1}{2}\,\langle \tilde{\beta}, \omega\otimes\omega\rangle - \varepsilon \\ &\quad + \liminf_{n\to\infty} \tfrac{1}{a_n n^2} \log \tilde{\mathbb{P}}\big\{(L^1,\, L^2) \in B_\omega^1 \times B_\varpi^2\big\}.\end{aligned}$$

The result follows once we prove that

$$\liminf_{n\to\infty} \tfrac{1}{a_n n^2} \log \tilde{\mathbb{P}}\big\{(L^1,\, L^2) \in B_\omega^1 \times B_\varpi^2\big\} = 0. \tag{3.5.22}$$

To conclude the proof, we use the upper bound (but now with the law $\mathbb{P}$ replaced by $\tilde{\mathbb{P}}$) and the LDP on the scale n, Theorem 3.5.1(ii), to prove (3.5.22).

Therefore, we have

$$\begin{aligned}\limsup_{n\to\infty} \tfrac{1}{a_n n^2} \log \tilde{\mathbb{P}}\big\{(L^1,\, L^2) \in (B^1_\omega \times B^2_\varpi)^c\big\} \\ \le \limsup_{n\to\infty} \tfrac{1}{n} \log \tilde{\mathbb{P}}\big\{(L^1,\, L^2) \in (B^1_\omega \times B^2_\varpi)^c\big\} \le - \inf_{(\tilde\omega,\,\tilde\varpi)\in\hat F} \tilde I_2(\tilde\omega,\, \tilde\varpi),\end{aligned}$$

$$\tilde I_2(\tilde\omega,\, \tilde\varpi) = \begin{cases} H(\,\tilde\omega \,\|\, \omega\,) & \text{if } \tilde\varpi = \tilde C\tilde\omega \otimes \tilde\omega, \\ \infty & \text{otherwise,}\end{cases}$$

where $\hat F = (B^1_\omega \times B^2_\varpi)^c$ and $(B^1_\omega \times B^2_\varpi)^c$ is the complement of the set $B^1_\omega \times B^2_\varpi$.

It remain for us to show that the infimum is positive. Suppose by contradiction there exists the sequence $(\omega_n,\, \varpi_n) \in \hat F$ such that $\tilde I_2(\tilde\omega,\, \tilde\varpi) \downarrow 0$. Then, because $\tilde I_2$ is good rate function with all its level sets are compact, and by lower semicontinuity of the map $(\tilde\omega,\, \tilde\varpi) \to \tilde I_2(\tilde\omega,\, \tilde\varpi)$, we can construct a limit point $(\tilde\varpi,\, \tilde\omega) \in \hat F$ with $\tilde I_2(\tilde\omega,\, \tilde\varpi) = 0$. This means $\tilde\omega = \omega$ and $\tilde\varpi = \tilde C\tilde\omega \otimes \tilde\omega = \varpi$, and hence, contradicting $(\tilde\omega,\, \tilde\varpi) \in \hat F$. ■

3.5.5 Large Deviations for Sub-Critical Coloured Random Graphs on the Scale n

For $\tilde f \in \mathcal{C}_1$ we recall from Subsection 3.5.3 the definition of the constant

$$U_{\tilde f} = \log \sum_{a\in\mathcal{X}} e^{\tilde f(a)} \mu(a).$$

Also recall for $\tilde g \in \mathcal{C}_2$ the definition of the symmetric function

$$\tilde h^{(2)}_n(a,b) = \log\Big(1 - p_n(a,b) + p_n(a,b) e^{\tilde g(a,b)/n a_n}\Big)^{-n}.$$

We use $\tilde f,\ \tilde g$ to define (for sufficiently large n) a new random graph as follows:

- Assign vertices V colours from $\mathcal{X}$ independently and identically according to the colour law $\widetilde\mu$ defined by

$$\widetilde\mu(a) = e^{\tilde f(a) - U_{\tilde f}} \mu(a).$$

- Given any two vertices $u, v \in V$, with u carrying colour a and v carrying colour b connect vertex u to vertex v with probability

$$\tilde{p}_n(a,b) = \frac{p_n(a,b)e^{\tilde{g}(a,b)/na_n}}{1-p_n(a,b)+p_n(a,b)e^{\tilde{g}(a,b)/na_n}}, \tag{3.5.23}$$

otherwise keep u and v disconnected.

For this new graph, observe $\widetilde{\mu}$ is a probability measure and further that

$$\tilde{p}_n(a,\, b) \in [0,\, 1],\, \forall a, b \in \mathcal{X}.$$

Denote by $\tilde{\mathbb{P}}$ the law of the new coloured random graph construct from $\widetilde{\mu}$ and $\tilde{p}$.

We recall from the previous chapter that $L^1_\Delta = \frac{1}{n}\sum_{u\in V}\delta_{(X(u),X(u))}$.

We note from the construction of the new graph that $\tilde{\mathbb{P}}$ is absolutely continuous with respect to $\mathbb{P}$, as for a coloured random graph X,

$$\begin{aligned}
\frac{d\tilde{\mathbb{P}}}{d\mathbb{P}}(X) &= \prod_{u\in V}\frac{\tilde{\mu}(X(u))}{\mu(X(u))}\prod_{(u,v)\in E}\frac{\tilde{p}_n(X(u),X(v))}{p_n(X(u),X(v))}\prod_{(v,u)\notin E}\frac{\tilde{q}_n(X(u),X(v))}{q_n(X(u),X(v))}\\
&= \prod_{u\in V}\frac{\tilde{\mu}(X(u))}{\mu(X(u))}\prod_{(u,v)\in E}\frac{\tilde{p}_n(X(u),X(v))}{p_n(X(u),X(v))}\times\frac{n-np_n(X(u),X(v))}{n-n\tilde{p}_n(X(u),X(v))}\prod_{(u,v)\in\mathcal{E}}\frac{n-n\tilde{p}_n(X(u),X(v))}{n-np_n(X(u),X(v))}\\
&= \prod_{u\in V}e^{\tilde{f}(X(u))-U_{\tilde{f}}}\prod_{(u,v)\in E}e^{\tilde{g}(X(u),X(v))/na_n}\prod_{(u,v)\in\mathcal{E}}e^{\tilde{h}_n^{(2)}(X(u),X(v))/n}\\
&= e^{n\langle L^1,\tilde{f}-U_{\tilde{f}}\rangle+n\langle\frac{1}{2}L^2,\tilde{g}\rangle+n\langle\frac{1}{2}L^1\otimes L^1,\tilde{h}_n^{(2)}\rangle-n\langle L^1_\Delta,\tilde{h}_n^{(2)}\rangle}.
\end{aligned}\tag{3.5.24}$$

Upper Bound in Theorems 3.5.2(ii)

We define for $(\omega, \varpi) \in \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}_*(\mathcal{X}\times\mathcal{X})$, the function $\hat{I}_4(\omega,\, \varpi)$ by

$$\hat{I}_4(\omega,\, \varpi) = \sup_{\tilde{f}\in\mathcal{C}_1}\Big\{\sum_{a\in\mathcal{X}}\big(\tilde{f}(a)-U_{\tilde{f}}\big)\omega(a)\Big\} \tag{3.5.25}$$

Lemma 3.5.9. *For each closed set $F \subset \tilde{\mathcal{M}}_*(\mathcal{X}\times\mathcal{X})$,*

$$\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\big\{(L^1,\, L^2)\in F\big\} \le -\inf_{(\omega,\,\varpi)\in F}\hat{I}_4(\omega,\, \varpi).$$

Proof. Fix $\tilde{f} \in \mathcal{C}_1$ and take $\tilde{g}(a,b) = 0$, $\forall a, b \in \mathcal{X}$. We observe that

$$\lim_{n\to\infty} \tilde{h}_n^{(2)}(a,b) = 0, \ \ \forall a, b \in \mathcal{X}.$$

Using (3.5.24) we obtain $\mathbb{E}\{e^{n\langle L^1, \tilde{f} - U_{\tilde{f}}\rangle}\} = \int d\tilde{\mathbb{P}} \le 1$. Therefore, we have

$$\limsup_{n\to\infty} \tfrac{1}{n} \log \mathbb{E}\Big\{e^{n\langle L^1, \tilde{f} - U_{\tilde{f}}\rangle}\Big\} \le 0. \tag{3.5.26}$$

Now fix $\varepsilon > 0$ and write $\hat{I}_4^{\varepsilon}(\omega, \varpi) := \min\{\hat{I}_4(\omega, \varpi), \varepsilon^{-1}\} - \varepsilon$.

We suppose $(\omega, \varpi) \in F$ and choose $\tilde{f} \in \mathcal{C}_1$ such that $\langle \tilde{f} - U_{\tilde{f}}, \omega\rangle \ge \hat{I}_4^{\varepsilon}(\omega, \varpi)$.

By finiteness of $\mathcal{X}$, we can find open neighbourhoods B_{ϖ}^2 and B_{ω}^1 of ϖ, ω such that

$$\inf_{\tilde{\omega}\in B_{\omega}^1} \{\langle \tilde{\omega}, \ \tilde{f} - U_{\tilde{f}}\rangle\} \ge \hat{I}_4^{\varepsilon}(\omega, \varpi) - \varepsilon.$$

Using Chebysheff's inequality and (3.5.26), we have that

$$\begin{aligned}
&\limsup_{n\to\infty} \tfrac{1}{n} \log \mathbb{P}\{(L^1, L^2) \in B_{\omega}^1 \times B_{\varpi}^2\} \\
&\qquad \le \limsup_{n\to\infty} \tfrac{1}{n} \log \mathbb{E}\Big\{e^{n\langle L^1, \tilde{f} - U_{\tilde{f}}\rangle}\Big\} - \hat{I}_4^{\varepsilon}(\omega, \varpi) + \varepsilon \\
&\qquad \le -\hat{I}_4^{\varepsilon}(\omega, \varpi) + \varepsilon.
\end{aligned} \tag{3.5.27}$$

By Lemma 2.4.9 we choose $N(\varepsilon) \in \mathbb{N}$ (with $\alpha = \varepsilon^{-1}$,) such that

$$\limsup_{n\to\infty} \tfrac{1}{n} \log \mathbb{P}\Big\{|E| > a_n n^2 N(\varepsilon)\Big\} \le -\varepsilon^{-1}.$$

We define for this N the set $K_{N(\varepsilon)}$ by

$$K_{N(\varepsilon)} = \big\{(\omega, \varpi) \in \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X}) : \|\varpi\| \le 2N(\varepsilon)\big\}.$$

The set $K_{N(\varepsilon)} \cap F$ is compact and therefore may be covered by finitely many sets

$$B_{\omega_r}^1 \times B_{\varpi_r}^2, r = 1, \ldots, m \text{ with } (\omega_r, \varpi_r) \in F \text{ for } r = 1, \ldots, m.$$

Hence, we have

$$\mathbb{P}\big\{(L^1,\, L^2) \in F\big\} \le \sum_{r=1}^{m} \mathbb{P}\big\{(L^1, L^2) \in B^1_{\omega_r} \times B^2_{\varpi_r}\big\} + \mathbb{P}\big\{(L^1, L^2) \notin K_{N(\varepsilon)}\big\}.$$

Now we use (3.5.27) to obtain, for all sufficiently small $\varepsilon > 0$,

$$\begin{aligned}\limsup_{n\to\infty} \tfrac{1}{n} \log \mathbb{P}\big\{(L^1, L^2) \in F\big\} &\le \max_{r=1}^{m} \limsup_{n\to\infty} \tfrac{1}{n} \log \mathbb{P}\big\{(L^1, L^2) \in B^1_{\omega_r} \times B^2_{\varpi_r}\big\} \vee -\varepsilon^{-1} \\ &\le - \inf_{(\omega,\varpi)\in F} \hat{I}^{\varepsilon}_4(\omega, \varpi) \vee -\varepsilon^{-1} + \varepsilon.\end{aligned}$$

Taking $\varepsilon \downarrow 0$ we have the desired statement. ■

By the variational characterization of the relative entropy we have

$$I_4(\omega,\, \varpi) = H(\omega \,\|\, \mu),$$

which ends the proof of the upper bound.

Lower Bounds Theorems 3.5.2(ii)

Lemma 3.5.10. *For every open set $O \subset \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X})$,*

$$\liminf_{n\to\infty} \tfrac{1}{n} \log \mathbb{P}\big\{(L^1, L^2) \in O\big\} \ge - \inf_{(\omega,\, \varpi)\in O} I_4(\omega,\, \varpi). \qquad (3.5.28)$$

Proof. Suppose $(\omega,\, \varpi) \in O$ with $\varpi \ll C\omega \otimes \omega$. We define the function $\tilde{f}_\omega \colon \mathcal{X} \to \mathbb{R}$ by

$$\tilde{f}_\omega(a) = \begin{cases} \log \frac{\omega(a)}{\mu(a)}, & \text{if } \omega(a) > 0, \\ 0, & \text{otherwise,} \end{cases}$$

and the symmetric function $g_\varpi \colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ by

$$g_\varpi(a,b) = \begin{cases} \log \frac{\varpi(a,b)}{C(a,b)\omega(a)\omega(b)}, & \text{if } \varpi(a,b) > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Set $\tilde{g}_{\tilde{\varpi}}(a,b) = na_n g_\varpi(a,b),\ \forall a,b \in \mathcal{X}$ and note by this choice of $\tilde{g}$ we have

$$\lim_{n\to\infty} \tilde{h}^{(2)}_n(a,b) = 0,\ \ \forall a,b \in \mathcal{X}.$$

We write $l := 0 \wedge \min_{a,b\in\mathcal{X}} \tilde{g}_{\tilde{\varpi}}(a,b)$.

Choose B^1_ω, B^2_ϖ open neighbourhoods of ω, ϖ such that $B^1_\omega \times B^2_\varpi \subset O$ and

$$\forall(\tilde\omega, \tilde\varpi) \in B^1_\omega \times B^2_\varpi,$$
$$\langle \omega,\, \tilde f_\omega\rangle + \tfrac{1}{2}\,\|\varpi\|\varepsilon - \varepsilon \le \langle \tilde\omega,\, \tilde f_\omega\rangle + \tfrac{1}{2}\,\|\tilde\varpi\|\varepsilon.$$

We use the probability measure $\tilde{\mathbb{P}}$ given by $\tilde g_\varpi$. We observe that the colour law is ω and the connection probabilities satisfy

$$a_n^{-1}\tilde p_n(a,b) \xrightarrow{n\uparrow\infty} \tilde C(a,b) := \tfrac{\varpi(a,b)}{\omega(a)\omega(b)}.$$

Therefore, using (3.5.24) we have that

$$\begin{aligned}
\mathbb{P}\Big\{(L^1,L^2)\in O\Big\} &\ge \tilde{\mathbb{E}}\Big\{\tfrac{d\mathbb{P}}{d\tilde{\mathbb{P}}}(X)\mathbb{1}_{\{(L^1,L^2)\in B^1_\omega\times B^2_\varpi\}}\Big\}\\
&= \tilde{\mathbb{E}}\Big\{e^{-n\langle L^1,\tilde f_\omega\rangle - n\langle \frac{1}{2}L^2,\tilde g\rangle - n\langle\frac{1}{2}L^1\otimes L^1,\tilde g\rangle + n\langle L^1_\Delta,\tilde h_n^{(2)}\rangle}\mathbb{1}_{\{(L^1,L^2)\in B^1_\omega\times B^2_\varpi\}}\Big\}\\
&\ge e^{-n\,\langle\omega,\tilde f_\omega\rangle - n\,\|\varpi\|\varepsilon/2 - n\,\varepsilon - a_n n^2 l/2 + o(1)} \times \tilde{\mathbb{P}}\Big\{(L^1,L^2)\in B^1_\omega\times B^2_\varpi\Big\}.
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
\liminf_{n\to\infty}\tfrac{1}{n}\log\mathbb{P}\big\{(L^1,L^2)\in O\big\} \ge& -\langle\omega,\,\tilde f_\omega\rangle - \tfrac{1}{2}\,\|\varpi\|\varepsilon - \varepsilon\\
&+ \liminf_{n\to\infty}\tfrac{1}{n}\log\tilde{\mathbb{P}}\big\{(L^1,L^2)\in B^1_\omega\times B^2_\varpi\big\}.
\end{aligned}$$

The result follows once we prove that

$$\liminf_{n\to\infty}\tfrac{1}{n}\log\tilde{\mathbb{P}}\big\{(L^1,L^2)\in B^1_\omega\times B^2_\varpi\big\} = 0. \tag{3.5.29}$$

We use the upper bound (but now with the law $\mathbb{P}$ replaced by $\tilde{\mathbb{P}}$) to prove (3.5.13).

Therefore, we have

$$\limsup_{n\to\infty}\tfrac{1}{n}\log\tilde{\mathbb{P}}\big\{(L^1,L^2)\in(B^1_\omega\times B^2_\varpi)^c\big\} \le -\inf_{(\tilde\omega,\tilde\varpi)\in\tilde F}\tilde I_4(\tilde\omega,\tilde\varpi),$$

where $\tilde F = (B^1_\omega\times B^2_\varpi)^c$.

It therefore remain for us to show that the infimum is positive.

To do this, we suppose for contradiction that there exists a sequence $(\tilde{\omega}_n, \tilde{\varpi}_n) \in \tilde{F}$ with $\tilde{I}_4(\tilde{\omega}_n, \tilde{\varpi}_n) \downarrow 0$. Then, since $\tilde{I}_4$ is a good rate function and its level sets are compact, and the mapping $(\tilde{\omega}, \tilde{\varpi}) \mapsto \tilde{I}_4(\tilde{\omega}, \tilde{\varpi})$ is lower semicontinuous, we can construct a limit point $(\tilde{\omega}, \tilde{\varpi}) \in \tilde{F}$ with $\tilde{I}_4(\tilde{\omega}, \tilde{\varpi}) = 0$. This implies $H(\tilde{\omega} \,\|\, \omega) = 0$ and hence $\tilde{\omega} = \omega$, which contradicts $(\tilde{\omega}, \tilde{\varpi}) \in \tilde{F}$.

■

3.5.6 Large Deviations for Sub-Critical Coloured Random Graphs on the Scale $a_n n^2$

Recall that for a given $\tilde{f} \in \mathcal{C}_1$ the constant

$$U_{\tilde{f}} = \log \sum_{a \in \mathcal{X}} e^{\tilde{f}(a)} \mu(a).$$

Further, we recall that for $\tilde{g} \in \mathcal{C}_2$ the function $\tilde{h}_n^{(1)} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is

$$\tilde{h}_n^{(1)}(a,b) = -\log \left[\left(1 - p_n(a,b) + p_n(a,b) e^{\tilde{g}(a,b)}\right)^{1/a_n} \right], \quad \text{for } a, b \in \mathcal{X}.$$

For $\tilde{f}$, $\tilde{g}$ define (for sufficiently large n) a new random graph in the following way:

- Assign to the n vertices in V colours from $\mathcal{X}$ independently and identically according to the colour law $\widetilde{\mu}$ defined by

$$\widetilde{\mu}(a) = e^{\tilde{f}(a) - U_{\tilde{f}}} \mu(a).$$

- Given any two vertices $u, v \in V$, with u carrying colour a and v carrying colour b connect vertex u to vertex v with probability

$$\tilde{p}_n(a,b) = \tfrac{p_n(a,b) e^{\tilde{g}(a,b)}}{1 - p_n(a,b) + p_n(a,b) e^{\tilde{g}(a,b)}}, \tag{3.5.30}$$

  otherwise keep u and v disconnected.

Observe the colour law $\tilde{\mu}$ is a probability measure and the connection probabilities

$$\tilde{p}_n(a,\, b) \in [0,\, 1], \; \forall a, b \in \mathcal{X}.$$

Denote by $\tilde{\mathbb{P}}$ the law of the coloured random graph obtained from $\widetilde{\mu}$ and $\tilde{p}_n$.

Recall that

$$L^2_\Delta = \tfrac{1}{n^2} \sum_{u\in V} \delta_{(X(u),X(u))}.$$

By construction $\tilde{\mathbb{P}}$ is absolutely continuous with respect to $\mathbb{P}$, as for a coloured random graph X,

$$\begin{aligned}
\frac{d\tilde{\mathbb{P}}}{d\mathbb{P}}(X) &= \prod_{u\in V} \tfrac{\tilde{\mu}(X(u))}{\mu(X(u))} \prod_{(u,v)\in E} \tfrac{\tilde{p}_n(X(u),X(v))}{p_n(X(u),X(v))} \prod_{(v,u)\notin E} \tfrac{n-n\tilde{p}_n(X(u),X(v))}{n-np_n(X(u),X(v))} \\
&= \prod_{u\in V} e^{\tilde{f}(X(u))-U_{\tilde{f}}} \prod_{(u,v)\in E} \tfrac{\tilde{p}_n(X(u),X(v))}{p_n(X(u),X(v))} \times \tfrac{n-n\tilde{p}_n(X(u),X(v))}{n-np_n(X(u),X(v))} \prod_{(u,v)\in \mathcal{E}} \tfrac{n-np_n(X(u),X(v))}{n-n\tilde{p}_n(X(u),X(v))} \\
&= \prod_{u\in V} e^{\tilde{f}(X(u))-U_{\tilde{f}}} \prod_{(u,v)\in E} e^{\tilde{g}(X(u),\,X(v))} \prod_{(u,v)\in \mathcal{E}} e^{a_n \tilde{h}_n^{(1)}(X(u),\,X(v))} \\
&= e^{n\langle L^1,\, \tilde{f}-U_{\tilde{f}}\rangle + a_n n^2 \langle \frac{1}{2} L^2,\, \tilde{g}\rangle + a_n n^2 \langle \frac{1}{2} L^1\otimes L^1,\, \tilde{h}_n^{(1)}\rangle - a_n n^2 \langle L^2_\Delta,\, \tilde{h}_n^{(1)}\rangle}. \qquad (3.5.31)
\end{aligned}$$

Upper Bound in Theorem 3.5.2(i). To begin we obtain the upper bound in a variational formulation. We recall that $na_n \to 0$ for subcritical coloured graphs and write $Z_n(f) := \frac{1}{na_n} U_{na_n f}$. Notice

$$Z(f) := \lim_{n\to\infty} Z_n(f) < \infty.$$

We define for $(\omega, \varpi) \in \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}_*(\mathcal{X}\times\mathcal{X})$ the rate function $\hat{I}_3$ by

$$\begin{aligned}
\hat{I}_3(\omega, \varpi) = \sup_{\tilde{g}\in \mathcal{C}_2} \Big\{ \sum_{a\in\mathcal{X}} (f(a) - Z(f))\omega(a) &+ \sum_{a,b\in\mathcal{X}} \tfrac{1}{2} \tilde{g}(a,b)\varpi(a,b) \\
&+ \sum_{a,b\in\mathcal{X}} \tfrac{1}{2}(1 - e^{\tilde{g}(a,b)}) C(a,b)\omega(a)\omega(b) \Big\}.
\end{aligned} \qquad (3.5.32)$$

Lemma 3.5.11. *For each closed set $F \subset \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}_*(\mathcal{X}\times\mathcal{X})$,*

$$\limsup_{n\to\infty} \tfrac{1}{a_n n^2} \log \mathbb{P}\big\{(L^1,\, L^2) \in F\big\} \le - \inf_{(\omega,\, \varpi)\in F} \hat{I}_3(\omega,\, \varpi).$$

Proof. Fix $f \in \mathcal{C}_1$. For any $\tilde{g} \in \mathcal{C}_2$ we define $\tilde{\beta}\colon \mathcal{X}\times\mathcal{X} \to \mathbb{R}$ by

$$\tilde{\beta}(a,b) = (1 - e^{\tilde{g}(a,b)}) C(a,b).$$

From Lemma 2.4.8 we note that we have $\lim_{n\to\infty} \tilde{h}_n^{(1)}(a,b) = \tilde{\beta}(a,b)$, $\forall a, b \in \mathcal{X}$.

We take $\tilde{f}(a) = na_n f(a)$, $\forall a \in \mathcal{X}$ and use (3.5.31) to obtain

$$\begin{aligned} e^{2na_n \max_{a\in\mathcal{X}} |\tilde{\beta}(a,a)|} &\geq \int e^{a_n n^2 \langle \frac{1}{2} L_\Delta^2, \tilde{h}_n^{(1)}\rangle} d\tilde{\mathbb{P}} \\ &= \mathbb{E}\Big\{ e^{a_n n^2 \langle L^1, f - Z_n(f)\rangle + a_n n^2 \langle \frac{1}{2} L^2, \tilde{g}\rangle + a_n n^2 \langle \frac{1}{2} L^1 \otimes L^1, \tilde{h}_n^{(1)}\rangle} \Big\}. \end{aligned}$$

Therefore, we have

$$\limsup_{n\to\infty} \frac{1}{a_n n^2} \log \mathbb{E}\Big\{ e^{a_n n^2 \langle L^1, f - Z_n(f)\rangle + a_n n^2 \langle \frac{1}{2} L^2, \tilde{g}\rangle + a_n n^2 \langle \frac{1}{2} L^1 \otimes L^1, \tilde{h}_n^{(1)}\rangle} \Big\} \leq 0. \quad (3.5.33)$$

Fix $\varepsilon > 0$ and take $\hat{I}_3^\varepsilon(\omega,\, \varpi) = \min\{\hat{I}_3(\omega,\, \varpi), \varepsilon^{-1}\} - \varepsilon$. Suppose $(\omega,\, \varpi) \in F$.

Choose $f \in \mathcal{C}_1$, $\tilde{g} \in \mathcal{C}_2$ such that $\langle \omega,\, f - Z(f)\rangle + \frac{1}{2}\langle \varpi, \tilde{g}\rangle + \frac{1}{2}\langle \omega \otimes \omega,\, \tilde{\beta}\rangle \geq \hat{I}_3^\varepsilon(\omega, \varpi)$.

By finiteness of $\mathcal{X}$, we can find open neighbourhoods B_ω^1, B_ϖ^2 of ω, ϖ such that

$$\inf_{\tilde{\omega}\in B_\omega^1,\, \tilde{\varpi}\in B_\varpi^2} \big\{ \langle \tilde{\omega},\, f - Z(f)\rangle + \langle \tfrac{1}{2}\tilde{g},\, \tilde{\varpi}\rangle + \langle \tfrac{1}{2}\tilde{\omega} \otimes \tilde{\omega},\, \tilde{\beta}\rangle \big\} \geq \hat{I}_3^\varepsilon(\omega,\, \varpi) - \varepsilon.$$

By Chebysheff's inequality and (3.5.33), we have that

$$\begin{aligned} &\limsup_{n\to\infty} \frac{1}{a_n n^2} \log \mathbb{P}\big\{ (L^1,\, L^2) \in B_\omega^1 \times B_\varpi^2 \big\} \\ &\leq \limsup_{n\to\infty} \frac{1}{a_n n^2} \log \mathbb{E}\Big\{ e^{a_n n^2 \langle L^1, \tilde{f} - Z_n(f)\rangle + a_n n^2 \langle \frac{1}{2} L^2, \tilde{g}\rangle + a_n n^2 \langle \frac{1}{2} L^1 \otimes L^1, \tilde{h}_n^{(1)}\rangle} \Big\} - \hat{I}_3^\varepsilon(\omega,\, \varpi) + \varepsilon \\ &\leq -\hat{I}_3^\varepsilon(\omega,\, \varpi) + \varepsilon. \end{aligned} \quad (3.5.34)$$

Using Lemma (2.4.9) with $\alpha = \varepsilon^{-1}$ we choose $N(\varepsilon) \in \mathbb{N}$ such that

$$\limsup_{n\to\infty} \frac{1}{a_n n^2} \log \mathbb{P}\Big\{ |E| > a_n n^2 N(\varepsilon) \Big\} \leq -\varepsilon^{-1}.$$

Define for this N, the set $K_{N(\varepsilon)}$ by

$$K_{N(\varepsilon)} = \big\{ (\omega,\, \varpi) \in \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X}) : \|\varpi\| \leq 2N(\varepsilon) \big\}.$$

Note $K_{N(\varepsilon)} \cap F$ is compact and therefore may be covered by finitely many sets

$$B_\omega^1 \times B_{\varpi_r}^2, r = 1, ..., m \text{ with } \varpi_r \in F \text{ for } r = 1, ..., m.$$

Hence, we have

$$\mathbb{P}\big\{(L^1,\,L^2)\in F\big\}\le\sum_{r=1}^{m}\mathbb{P}\big\{(L^1,\,L^2)\in B^1_{\omega_r}\times B^2_{\varpi_r}\big\}+\mathbb{P}\big\{(L^1,\,L^2)\notin K_{N(\varepsilon)}\big\}.$$

Using (3.5.34) for small enough $\varepsilon>0$, we obtain

$$\begin{aligned}\limsup_{n\to\infty}\tfrac{1}{a_n n^2}\log\mathbb{P}\big\{(L^1,L^2)\in F\big\}&\\ &\le\max_{r=1}^{m}\limsup_{n\to\infty}\tfrac{1}{a_n n^2}\log\mathbb{P}\big\{(L^1,\,L^2)\in B^1_{\omega_r}\times B^2_{\varpi_r}\big\}\vee-\varepsilon^{-1}\\ &\le-\hat{I}_3^{\varepsilon}(\omega,\,\varpi)\vee-\varepsilon^{-1}+\varepsilon.\end{aligned}$$

Taking $\varepsilon\downarrow 0$ we have the required statement.

■

We identify the rate function by solving the variational problem in the right side of equation (3.5.32).

Lemma 3.5.12. *For any $(\omega,\,\varpi)\in\mathcal{M}(\mathcal{X})\times\tilde{\mathcal{M}}_*(\mathcal{X}\times\mathcal{X})$we have*

(i) $\hat{I}_3(\omega,\,\varpi)=I_3(\omega,\,\varpi)$ and (ii) I_3 is good rate function.

Proof. (i) Suppose $\omega\in\mathcal{M}(\mathcal{X})$ is not equal μ. Define the function f by

$$f(a)=K\log(|\omega(a)-\mu(a)|+1),\ \text{ for } a\in\mathcal{X} \text{ and } K\in\mathbb{R}.$$

Set $\tilde{g}(a,b)=0$ for all $a,b\in\mathcal{X}$ in (3.5.32) and note that by the choice of f,

$$\begin{aligned}&\sum_{a\in\mathcal{X}}(f(a)-Z(f))\omega(a)+\sum_{a,b\in\mathcal{X}}\tfrac12\tilde{g}(a,b)\varpi(a,b)\sum_{a,b\in\mathcal{X}}\tfrac12(1-e^{\tilde{g}(a,b)})C(a,b)\omega(a)\omega(b)\\ &\qquad\ge K\Big(\sum_{a\in\mathcal{X}}\log(|\omega(a)-\mu(a)|+1)\omega(a)-\max_{a}|\omega(a)-\mu(a)|-1\Big)\xrightarrow{|K|\uparrow\infty}\infty,\end{aligned}$$

where the sign of $|K|$ is such that last expression always stays positive.
Suppose $\varpi\not\ll C\omega\otimes\omega$. Then there exists $a_0,\,b_0\in\mathcal{X}$ with

$$C(a_0,b_0)\omega(a_0)\omega(b_0)=0\qquad\text{and}\qquad\varpi(a_0,b_0)>0.$$

For this (a_0, b_0) we define the symmetric function $\tilde{g}$ by

$$\tilde{g}(a,b) = \log(K(\mathbb{1}_{(a_0,b_0)}(a,b) + \mathbb{1}_{(b_0,a_0)}(a,b)) + 1), \quad \text{for } a,b \in \mathcal{X} \text{ and } K > 0.$$

Considering our $\tilde{g}$ in (3.5.17) we have

$$\sum_{a,b\in\mathcal{X}} \tfrac{1}{2}\tilde{g}(a,b)\varpi(a,b) + \sum_{a,b\in\mathcal{X}} \tfrac{1}{2}(1 - e^{\tilde{g}(a,b)}C(a,b)\omega(a)\omega(b)$$
$$= \log(K+1)(\varpi(a_0,b_0)) \xrightarrow{K\uparrow\infty} \infty.$$

Suppose that $\varpi \ll C\omega \otimes \omega$. Then, we have

$$\hat{I}(\omega,\varpi) \geq \tfrac{1}{2} \sup_{g\in\mathcal{C}_2} \Big\{ \sum_{a,b\in\mathcal{X}} g(a,b)\varpi(a,b) - \sum_{a,b\in\mathcal{X}} e^{g(a,b)}C(a,b)\omega(a)\omega(b) \Big\}$$
$$+ \tfrac{1}{2} \sum_{a,b\in\mathcal{X}} C(a,b)\omega(a)\omega(b).$$

Again by the substitution $h = e^{g} \frac{C\omega\otimes\omega}{\varpi}$ and $\sup_{x>0} \log x - x = -1$ we have

$$\sup_{g\in\mathcal{C}_2} \Big\{ \sum_{a,b\in\mathcal{X}} g(a,b)\varpi(a,b) - \sum_{a,b\in\mathcal{X}} e^{g(a,b)}C(a,b)\omega(a)\omega(b) \Big\}$$
$$= \sup_{\substack{h\in\mathcal{C}_2\\ h\geq 0}} \sum_{a,b\in\mathcal{X}} \Big[\log\Big(h(a,b)\frac{\varpi(a,b)}{C(a,b)\omega(a)\omega(b)} \Big) - h(a,b) \Big] \varpi(a,b)$$
$$= \sup_{\substack{h\in\mathcal{C}_2\\ h\geq 0}} \sum_{a,b\in\mathcal{X}} \big(\log h(a,b) - h(a,b) \big) \varpi(a,b) + \sum_{a,b\in\mathcal{X}} \log\Big(\frac{\varpi(a,b)}{C(a,b)\omega(a)\omega(b)} \Big) \varpi(a,b)$$
$$= -\|\varpi\| + H(\varpi \,\|\, C\omega \otimes \omega).$$

This yields that

$$\hat{I}_3(\omega,\, \varpi) = I_3(\omega,\, \varpi).$$

(ii) This follows from the proof of Lemma 2.4.13. We recall from the proof of Lemma 2.4.13 that $\mathfrak{H}_C$ is a rate function and that for all $\alpha < \infty$, its level sets

$$\{(\omega,\varpi) \in \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}(\mathcal{X}\times\mathcal{X}) \colon\ \mathfrak{H}_C(\varpi \,\|\, \omega) \leq \alpha\}$$

are bounded, closed set and so are compact. This means I_3 is good rate function.

■

Lower Bound in Theorem 3.5.2(i). We use the upper bound (but now with the law $\mathbb{P}$ replaced by $\tilde{\mathbb{P}}$) to establish the lower bound for some open set

$$O \subset \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X}).$$

Lemma 3.5.13. *For every open set $O \subset \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X})$,*

$$\liminf_{n\to\infty} \frac{1}{a_n n^2} \log \mathbb{P}\big\{(L^1,\, L^2) \in O\big\} \geq - \inf_{(\tilde{\omega},\, \tilde{\varpi})\in O} I_3(\omega,\, \varpi). \tag{3.5.35}$$

Proof. Suppose $(\omega,\, \varpi) \in O$ with $\varpi \ll C\omega \otimes \omega$ and $\omega = \mu$. Take

$$\tilde{f}(a) = 0, \;\; \forall a \in \mathcal{X}.$$

Define the symmetric function $\tilde{g}_{\varpi}\colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ by

$$\tilde{g}_{\varpi}(a, b) = \begin{cases} \log \frac{\varpi(a,b)}{C(a,b)\omega(a)\omega(b)}, & \text{if } \varpi(a, b) > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Recall that $\tilde{h}_n^{(1)}(a, b) = -\log\Big[1 - p_n(a, b) + p_n(a, b)e^{\tilde{g}_{\varpi}(a,b)}\Big]^{1/a_n}$, for $a, b \in \mathcal{X}$.

Define the symmetric function $\tilde{\beta}_{\varpi}(a, b)$ by

$$\tilde{\beta}_{\varpi}(a, b) := \lim_{n\to\infty} \tilde{h}_n^{(1)}(a, b) = C(a, b)(1 - e^{g_{\varpi}(a,b)}).$$

Choose $B^1_{\omega}, B^2_{\varpi}$ open neighbourhoods of ω, ϖ such that $B^1_{\omega} \times B^2_{\varpi} \subset O$ and

$$\forall (\tilde{\omega}, \tilde{\varpi}) \in B^1_{\omega} \times B^2_{\varpi},$$
$$\langle \varpi,\, \tilde{g}_{\varpi}\rangle + \langle \omega \otimes \omega,\, \tilde{\beta}_{\varpi}\rangle - \varepsilon \leq \langle \tilde{\varpi},\, \tilde{g}_{\varpi}\rangle + \langle \tilde{\omega} \otimes \tilde{\omega},\, \tilde{\beta}_{\varpi}\rangle.$$

We note that, the coloured random graph obtained from the function $\tilde{g}_{\varpi}$ has colour law ω and connection probabilities satisfying

$$a_n^{-1}\tilde{p}_n(a, b) \xrightarrow{n\uparrow\infty} C(a, b) := \frac{\varpi(a,b)}{(\omega(a)\omega(b))}.$$

Write

$$m := 0 \wedge \min_{a\in\mathcal{X}} \tilde{\beta}_{\varpi}(a, a).$$

Therefore, by (3.5.31) we have

$$\begin{aligned}
\mathbb{P}\Big\{(L^1, L^2)\in O\Big\} &\geq \tilde{\mathbb{E}}\Big\{\tfrac{d\mathbb{P}}{d\tilde{\mathbb{P}}}(X)\mathbb{1}_{\{(L^1,\,L^2)\in B^1_\omega\times B^2_\varpi\}}\Big\}\\
&= \tilde{\mathbb{E}}\Big\{e^{-a_nn^2\,\langle\frac{1}{2}\,L^2,\tilde{g}_\varpi\rangle-a_nn^2\,\langle\frac{1}{2}\,L^1\otimes L^1,\tilde{h}_n^{(1)}\rangle+a_nn^2\,\langle \mathbf{L}^2_\Delta,\tilde{h}_n^{(1)}\rangle}\times\mathbb{1}_{\{(L^1,\,L^2)\in B^1_\omega\times B^2_\varpi\}}\Big\}\\
&\geq e^{-a_nn^2\langle\frac{1}{2}\,\varpi,\tilde{g}_\varpi\rangle-a_nn^2\langle\frac{1}{2}\,\omega\otimes\omega,\tilde{\beta}\rangle+a_nm/4-\frac{1}{2}\,a_nn^2\varepsilon}\times\tilde{\mathbb{P}}\Big\{(L^1,\,L^2)\in B^1_\omega\times B^2_\varpi\Big\}.
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
\liminf_{n\to\infty}\tfrac{1}{a_nn^2}\log\mathbb{P}\big\{(L^1,\,L^2)\in O\big\} \geq -\tfrac{1}{2}\,\langle\varpi,\,\tilde{g}\rangle-\tfrac{1}{2}\,\langle\omega\otimes\omega,\,\tilde{\beta}\rangle-\varepsilon\\
+\liminf_{n\to\infty}\tfrac{1}{a_nn^2}\log\tilde{\mathbb{P}}\big\{(L^1,\,L^2)\in B^1_\omega\times B^2_\varpi\big\}.
\end{aligned}$$

The result follows once we prove that

$$\liminf_{n\to\infty}\tfrac{1}{a_nn^2}\log\tilde{\mathbb{P}}\big\{(L^1,\,L^2)\in B^1_\omega\times B^2_\varpi\big\}=0. \tag{3.5.36}$$

We use the upper bound (but now with the law $\mathbb{P}$ replaced by $\tilde{\mathbb{P}}$) to prove (3.5.36). Therefore, we have

$$\limsup_{n\to\infty}\tfrac{1}{a_nn^2}\log\tilde{\mathbb{P}}\big\{(L^1,\,L^2)\in(B^1_\omega\times B^2_\varpi)^c\big\}\leq-\inf_{(\tilde{\omega},\,\tilde{\varpi})\in\hat{F}}\tilde{I}_3(\tilde{\omega},\,\tilde{\varpi}),$$

$$\tilde{I}_3(\tilde{\omega},\,\tilde{\varpi})=\begin{cases}\mathfrak{H}_C(\tilde{\varpi}\,\|\,\omega) & \text{if } \tilde{\omega}=\omega,\\ \infty & \text{otherwise.}\end{cases}$$

where $\hat{F}=(B^1_\omega\times B^2_\varpi)^c$ and $(B^1_\omega\times B^2_\varpi)^c$ is the complement of the set $B^1_\omega\times B^2_\varpi$.

It remain for us to show that the infimum is positive.

To show this, we suppose by contradiction there exists the sequence $(\omega_n,\,\varpi_n)\in\hat{F}$ such that $\tilde{I}_3(\tilde{\omega},\,\tilde{\varpi})\downarrow 0$. Then, because $\tilde{I}_3$ is good rate function with all its level sets compact, and by lower semicontinuity of the mapping $(\tilde{\omega},\,\tilde{\varpi})\to\tilde{I}_3(\tilde{\omega},\,\tilde{\varpi})$, we can construct a limit point $(\tilde{\varpi},\,\tilde{\omega})\in\hat{F}$ with $\tilde{I}_3(\tilde{\omega},\,\tilde{\varpi})=0$. This means $\tilde{\omega}=\omega$ and $\tilde{\varpi}=\tilde{C}\tilde{\omega}\otimes\tilde{\omega}=\varpi$, and hence, contradicting $(\tilde{\omega},\,\tilde{\varpi})\in\hat{F}$. ∎

3.5.7 Derivation of the Main Theorems

Lemma 3.5.14. *Suppose that X is a coloured random graph with colour law $\mu\colon \mathcal{X} \to (0,1]$ and connection probabilities $p_n : \mathcal{X} \times \mathcal{X} \to [0,1]$ such that $a_n^{-1}p_n(a,b) \to C(a,b)$ for some sequence (a_n) with $a_n n \to 0$ or $a_n n \to 1$ or $a_n n \to \infty$ and $C : \mathcal{X} \times \mathcal{X} \to [0,\infty)$. Then, for any $\varepsilon > 0$ we have*

$$\lim_{n\to\infty} \mathbb{P}\big\{ \sup_{a\in\mathcal{X}} |L^1(a) - \mu(a)| \geq \varepsilon \big\} = 0$$

and

$$\lim_{n\to\infty} \mathbb{P}\big\{ \sup_{a,b\in\mathcal{X}} |L^2(a,b) - \mu(a)C(a,b)\mu(b)| \geq \varepsilon \big\} = 0.$$

From Theorem 2.4.4, Theorem 3.5.1 and Theorem 3.5.2 we prove this lemma.

To begin, we define a closed set

$$F_1 = \big\{(\omega,\varpi) \in \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}_*(\mathcal{X}\times\mathcal{X})\colon \sup_{a,b\in\mathcal{X}} |\varpi(a,b) - \mu(a)C(a,b)\mu(b)| \geq \varepsilon\big\}.$$

We observe that in the sparse case (when $na_n \to 1$), by Theorem 2.4.4,

$$\limsup_{n\to\infty} \tfrac{1}{n} \log \mathbb{P}\Big\{(L^1, L^2) \in F_1\Big\} \leq - \inf_{(\omega,\varpi)\in F_1} I(\omega,\varpi). \tag{3.5.37}$$

We show by contradiction that the right hand side of (3.5.37) is negative. For this purpose suppose that there exists sequence (ω_n,ϖ_n) in F_1 such that $I(\omega_n,\varpi_n) \downarrow 0$. Then, because I is a good rate function and its level sets are compact, and by lower semicontinuity of the mapping $(\omega,\varpi) \mapsto I(\omega,\varpi)$, there is a limit point $(\omega,\varpi) \in F_1$ with $I(\omega,\varpi) = 0$. By Lemma 2.4.13, we have $H(\omega \,\|\, \mu) = 0$ and

$$\mathfrak{H}_C(\varpi \,\|\, \omega) = 0.$$

This implies $\omega(a) = \mu(a)$, and $\varpi(a,b) = C(a,b)\omega(a)\omega(b)$, for $a,b \in \mathcal{X}$ which contradicts $(\omega,\varpi) \in F_1$. Hence $\lim\limits_{n\to\infty} \mathbb{P}\big\{(L^1,L^2) \in F_1\big\} = 0$, as required.

For the subcritical case we can argue similarly with the LDP on the scale $a_n n^2$ with rate function I_3.

The first statement of Lemma 3.5.14 follows similarly using the set

$$F_2 = \big\{(\omega, \varpi) \in \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X}) \colon \sup_{a \in \mathcal{X}} |\omega(a) - \mu(a)| \geq \varepsilon\}$$

and the LDP of Theorem 2.4.4 in the sparse case, and the LDP on the scale n with rate function I_4 in the subcritical case.

Finally, in the supercritical case, an analogous argument can be carried out using $F = F_1 \cup F_2$ and the LDP on the scale n with rate function I_2.

We now compute the distribution $P_n \colon \mathcal{G}_n(\mathcal{X}) \to [0,1]$ of X,

$$\begin{aligned} P_n(x) &= \prod_{u \in V} \mu(x(u)) \prod_{(u,v) \in E} p_n(x(u), x(v)) \prod_{(u,v) \notin E} \big(1 - p_n(x(u), x(v))\big) \\ &= \prod_{u \in V} \mu(x(u)) \prod_{(u,v) \in E} \tfrac{p_n(x(u),x(v))}{1 - p_n(x(u),x(v))} \prod_{(u,v) \in \mathcal{E}} \big(1 - p_n(x(u), x(v))\big). \end{aligned}$$

Therefore, we have in the case of Theorem 3.4.1

$$\begin{aligned} -\tfrac{1}{a_n n^2 \log n} \log P_n(x) = \langle L^1, -\tfrac{\log \mu}{a_n\, n\, \log n}\rangle + \tfrac{1}{2} \langle L^2, -\tfrac{\log(p_n/(1-p_n))}{\log n}\rangle \\ + \tfrac{1}{2} \langle L^1 \otimes L^1, -\tfrac{\log(1-p_n)}{a_n \log n}\rangle + \tfrac{1}{2} \langle L^1_\Delta, -\tfrac{\log(1-p_n)}{a_n\, n\, \log n}\rangle. \end{aligned}$$

In the case of Theorem 3.4.2 we have

$$\begin{aligned} -\tfrac{1}{n} \log P(x) = \langle L^1, -\log \mu\rangle + \tfrac{1}{2} \langle L^2, -\tfrac{\log(p_n/(1-p_n))}{\log n}\rangle \\ + \tfrac{1}{2} \langle L^1 \otimes L^1, -n \log(1 - p_n)\rangle + \tfrac{1}{2} \langle L^1_\Delta, -\log(1 - p_n)\rangle. \end{aligned}$$

Now in the first case the integrands

$$\tfrac{-\log \mu}{a_n\, n\, \log n}, \quad \tfrac{-\log(1-p_n)}{a_n \log n} \text{ and } \tfrac{-\log(1-p_n)}{a_n\, n\, \log n}$$

all converge to zero, while

$$\tfrac{-\log(p_n/(1-p_n))}{\log n} \longrightarrow \mathbb{1}.$$

Hence Theorem 3.4.1 follows from Theorem 3.5.14.

In the second case the both integrand $-\log(1-p_n)$ and $-n\ \log(1-p_n)$ converges to zero. Therefore, Theorem 3.4.2 follows from Lemma 3.5.14.

Index of Notation

Symbol	Meaning
$\mathcal{X}$	Finite set of alphabet endowed with the discrete topology.
$\mathcal{M}(\mathcal{X})$	Set of probability measures on $\mathcal{X}$ equipped with the weak topology
$\tilde{\mathcal{M}}(\mathcal{X})$	Set of finite measures on $\mathcal{X}$ equipped with the weak topology.
$\tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X})$	Subspace of symmetric members of $\tilde{\mathcal{M}}(\mathcal{X} \times \mathcal{X})$.
$\mathcal{N}(\mathcal{X})$	Set of counting measures on $\mathcal{X}$ equipped with the discrete topology.
$\mathcal{C}_1$	Space of functions on $\mathcal{X}$.
$\mathcal{C}_2$	Space of symmetric functions on $\mathcal{X} \times \mathcal{X}$.
$\langle f\,,\,\omega \rangle$	Expectation of the function f with respect to the measure ω.
$\langle \nu \rangle$	Mean of the probability distribution ν.
$H(\omega \,\|\, \mu)$	Relative entropy of the probability vector ω with respect to μ.
$\|\ell\|$	Total mass of the measure ℓ.
d	Metric of total variation.
$\mathbb{P}$	Probability law of random graphs.
$\mathbb{Q}$	Offspring kernel.
$\Gamma(\cdot)$	Gamma function.
$\mathcal{S}(\nu)$	Support of the measure ν.
$m(a, c)$	Multiplicity of the symbol a in $c = (a_1, \ldots, a_n)$.
$A(a, b)$	Expected number of offspring of type a of a vertex of type b.
$P_n(x)$	Distribution of random graph (including trees) x on n vertices.
π	Eigenvector (normalized to a probability vector) corresponding to 1.

Bibliography

[Bar85] A.R. BARRON. The strong ergodic theorem for densities: Generalized Shannon-McMillan-Breiman Theorem. *The Annals of Probab.*, **13**, No.4, pp. 1292-1303 (1985).

[BBL04] S. BOUCHERON, O. BOUSQUET and G. LUGOSI. Concentration inequalities. In: *Advanced lectures in machine learning*, Eds: O. Bousquet, U. v. Luxburg and G. Rötsch. Springer (2004) 208-240.

[BGL01] S. BOUCHERON, F. GAMBOA and C. LEONARD. "Bins and Balls":Large deviations of the empirical occupany process, *Annals of Applied Probab.* Vol. 12 No 2, 1-30 (2000).

[Bi04] J.D. BIGGINS. Large deviations for mixtures. *El. Comm. Probab.* **9** 60–71 (2004).

[BP03] J.D. BIGGINS and D.B. PENMAN. Large deviations in randomly coloured random graphs. Preprint (2003).

[Bre57] L. BREIMAN. A correction to the individual ergodic theorem of information, Ann. Math. Stat.,**13**, pp. 809-810 (1960).

[CP03] C. CANNINGS and D.B. PENMAN. Models of random graphs and their applications *Handbook of Statistics 21. Stochastic Processes: Modeling and Simulation.* Eds: D.N. Shanbhag and C.R. Rao. Elsevier (2003) 51-91.

[CT91] T.M. COVER and J.A. THOMAS. Elements of Information Theory. Wiley Series in Telecommunications (1991).

[DA06b] K. DOKU-AMPONSAH. Asymptotic Equipartition Properties for Simple Hierarchical and Networked Structures. Preprint (2007).

[DK02] A. DEMBO and I. KONTOYIANNIS. Source Coding, Large deviations and Approximate Pattern. Invited paper in IEEE Transaction on information Theory, 48(6):1590-1615, June (2002).

[DM06a] K. DOKU-AMPONSAH and P. MÖRTERS. Large deviation principle the empirical measures of coloured random graphs. *Mathematic arXiv: math.PR/0607545v1 21 Jul 2006.*

[DMS03] A. DEMBO, P. MÖRTERS and S. SHEFFIELD. Large deviations of Markov chains indexed by random trees. *Ann. Inst. Henri Poincaré: Probab.et Stat.41,* (2005) 971-996.

[DZ98] A. DEMBO and O. ZEITOUNI. Large deviations techniques and applications. Springer, New York, (1998).

[Fel67] W. FELLER. An introduction to probability theory and its applications. Vol. I, Wiley, New York. Third edition, (1967).

[KA02] M. KIMMEL and D .E. AXELROD. Branching Processes with Biology. Springer, New York, (2002).

[McM53] B. MCMILLAN. The basic theorems of information theory. *Annal of Math. Statistics*, **24**, pp. 196-219(1953).

[Mo71] C.J. MODE Multitype Branching Processes Theory and Applications. American Elsevier, New York, (1971).

[Moy61] S.C. MOY. Generalization of Shannon-McMillan-Breiman theorem Pacific J. Math., **11**,pp. 705-7149(1961).

[New00] M. E. NEWMAN. Random graphs as models of networks. *http://arxiv.org/abs/cond-mat/0202208.*

[OC98] N. O'CONNELL. Some large deviation results for sparse random graphs. *Probab. Theory Relat. Fields* **110** 277–285 (1998).

[OS02] P. OLOFSSON and C.A. SHAW. Exact sampling formulas for multitype Galton-Watson processes. *Journal of Mathematical Biology* (2002) **45**, 279-293.

[Pe98] D.B. PENMAN. Random graphs with correlation structure. PhD Thesis, Sheffield 1998.

[Sha48] C.E. SHANNON. A mathematical theory of communication. Bell System Tech. J., **27**, pp.379-423, 623-656 (1948).